

**PCT**WORLD INTELLECTUAL PRO  
International

C3

INTERNATIONAL APPLICATION PUBLISHED UNDER

WO 9605315A2

<b>(51) International Patent Classification <sup>6</sup>:</b> C12N 15/51, C07K 14/18, G01N 33/569, A61K 39/29, C12Q 1/68, 1/70, C07K 16/10		<b>A2</b>	<b>(11) International Publication Number:</b> WO 96/05315 <b>(43) International Publication Date:</b> 22 February 1996 (22.02.96)
<b>(21) International Application Number:</b> PCT/US95/10398 <b>(22) International Filing Date:</b> 15 August 1995 (15.08.95) <b>(30) Priority Data:</b> 08/290,665 15 August 1994 (15.08.94) US <b>(71) Applicant:</b> THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY, DE- PARTMENT OF HEALTH AND HUMAN SERVICES, Of- fice of Technology Transfer National Institutes of Health [US/US]; Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US). <b>(72) Inventors:</b> BUKH, Jens; 5805 Sonoma Road, Bethesda, MD 20817 (US). MILLER, Roger, H.; 15504 White Willow Lane, Rockville, MD 20853 (US). PURCELL, Robert, H.; 17517 White Grounds Road, Boyds, MD 20841 (US). <b>(74) Agent:</b> FEILER, William, S.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).		<b>(81) Designated States:</b> AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
<b>(54) Title:</b> NUCLEOTIDE AND AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND CORE GENES OF HEPATITIS C VIRUS			
<b>(57) Abstract</b>  The nucleotide and deduced amino acid sequences of cDNAs encoding the envelope (1) genes and core genes of isolates of hepatitis C virus (HCV) are disclosed. The invention relates to the oligonucleotides, peptides and recombinant envelope (1) and core proteins derived from these sequences and their use in diagnostic methods and vaccines.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

- 1 -

Title of the InventionNUCLEOTIDE AND AMINO ACID SEQUENCES OF THE ENVELOPE 1  
AND CORE GENS OF HEPATITIS C VIRUS

5

The present application is a continuation-in-part of pending U.S. Application Serial No. 08/086/428, filed on June 29, 1993.

10

Field Of Invention

The present invention is in the field of hepatitis virology. The invention relates to the complete nucleotide and deduced amino acid sequences of the envelope 1 (E1) and core genes of hepatitis C virus (HCV) isolates from around the world and the grouping of these isolates into fourteen distinct HCV genotypes. More specifically, this invention relates to oligonucleotides, peptides and recombinant proteins derived from the envelope 1 and core gene sequences of these isolates of hepatitis C virus and to diagnostic methods and vaccines which employ these reagents.

20

Background Of Invention

Hepatitis C, originally called non-A, non-B hepatitis, was first described in 1975 as a disease serologically distinct from hepatitis A and hepatitis B (Feinstone, S.M. et al. (1975) N. Engl. J. Med. 292:767-770). Although hepatitis C was (and is) the leading type of transfusion-associated hepatitis as well as an important part of community-acquired hepatitis, little progress was made in understanding the disease until the recent identification of hepatitis C virus (HCV) as the causative agent of hepatitis C via the cloning and sequencing of the HCV genome (Choo, A.L. et al. (1989) Science 288:359-362). The sequence information generated by this study resulted

30

35

- 2 -

° in the characterization of HCV as a small, enveloped, positive-stranded RNA virus and led to the demonstration that HCV is a major cause of both acute and chronic hepatitis worldwide (Weiner, A.J. et al. (1990) Lancet 335:1-3). These observations, combined with studies showing that over 50% of acute cases of hepatitis C progress to chronicity with 20% of these resulting in cirrhosis and an undetermined proportion progressing to liver cancer, have led to tremendous efforts by investigators within the hepatitis C field to develop diagnostic assays and vaccines which can detect and prevent hepatitis C infection.

The cloning and sequencing of the HCV genome by Choo et al. (1989) has permitted the development of serologic tests which can detect HCV or antibody to HCV (Kuo, G. et al. (1989) Science 244:362-364). In addition, the work of Choo et al. has also allowed the development of methods for detecting HCV infection via amplification of HCV RNA sequences by reverse transcription and cDNA polymerase chain reaction (RT-PCR) using primers derived from the HCV genomic sequence (Weiner, A.J. et al.). However, although the development of these diagnostic methods has resulted in improved diagnosis of HCV infection, only approximately 60% of cases of hepatitis C are associated with a factor identified as contributing to transmission of HCV (Alter, M.J. et al. (1989) JAMA 262:1201-1205). This observation suggests that effective control of hepatitis C transmission is likely to occur only via universal pediatric vaccination as has been initiated recently for hepatitis B virus. Unfortunately, attempts to date to protect chimpanzees from hepatitis C infection via administration of recombinant vaccines have had only limited success. Moreover, the apparent genetic heterogeneity of HCV, as indicated by the recent assignment of all available HCV isolates to one of four genotypes, I-IV (Okamoto, H. et al. (1992) J. Gen. Virol; 73:673-679),



- 3 -

- ° presents additional hurdles which must be overcome in order to develop accurate and effective diagnostic assays and vaccines.

For example, one possible obstacle to the development of effective hepatitis C vaccines would arise if the observed genetic heterogeneity of HCV reflects serologic heterogeneity. In such a case, the most genetically diverse strains of HCV may then represent different serotypes of HCV with the result being that infection with one strain may not protect against infection with another. Indeed, the inability of one strain to protect against infection with another strain was recently noted by both Farci et al. (Farci, P. et al. (1992) Science 258:135-140) and Prince et al. (Prince, A.M. et al. (1992) J. Infect. Dis. 165:438-443), each of whom presented evidence that while infection with one strain of HCV does modify the degree of the hepatitis C associated with the reinfection, it does not protect against reinfection with a closely related strain. The genetic heterogeneity among different HCV strains also increases the difficulty encountered in developing RT-PCR assays to detect HCV infection since such heterogeneity often results in false-negative results because of primer and template mismatch. In addition, currently used serologic tests for detection of HCV or for detection of antibody to HCV are not sufficiently well developed to detect all of the HCV genotypes which might exist in a given blood sample. Finally, in terms of choosing the proper treatment modality to combat hepatitis infection, the inability of presently available serologic assays to distinguish among the various genotypes of HCV represents a significant shortcoming in that recent reports suggest that an HCV-infected patient's response to therapy might be related to the genotype of the infectious virus (Yoshioka, K. et al. (1992) Hepatology 16:293-299; Kanai, K. et al. (1992) Lancet 339:1543; Lan, J.Y.N. et al. (1992) Hepatology 16:209A). Indeed, the data

- 4 -

presented in the above studies suggest that the closely related genotypes I and II are less responsive to interferon therapy than are the closely related genotypes III and IV. Moreover, preliminary data by Pozzato et al. (Pozzato, G. et al. (1991) Lancet 338:509) suggests that different genotypes may be associated with different types or degrees of clinical disease. Taken together, these studies suggest that before effective vaccines against HCV infection can be developed, and indeed, before more accurate and effective methods for diagnosis and treatment of HCV infection can be produced, one must obtain a greater knowledge about the genetic and serologic diversity of HCV isolates.

In a recent attempt to gain an understanding of the extent of genetic heterogeneity among HCV strains, Bukh et al. carried out a detailed analysis of HCV isolates via the use of PCR technology to amplify different regions of the HCV genome (Bukh, J. et al. (1992a) Proc. Natl. Acad. Sci. 89:187-191). Following PCR amplification, the 5'-noncoding (5' NC) portion of the genomes of various HCV isolates were sequenced and it was found that primer pairs designed from conserved regions of the 5' NC region of the HCV genome were more sensitive for detecting the presence of HCV than were primer pairs representing other portions of the genome (Bukh, J. et al. (1992b) Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946). In addition, the authors noted that although many of the HCV isolates examined could be classified into the four genotypes described by Okamoto et al. (1992), other previously undescribed genotypes emerged based on genetic heterogeneity observed in the 5' NC region of the various isolates. One of the most prominent of these newly noted genotypes comprised a group of related viruses that contained the most genetically divergent 5' NC regions of those studied. This group of viruses, tentatively classified as a fifth genotype, are very similar to strains recently described by others (Cha, T.-A

- 5 -

° et al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:7144-7148;  
Chan, S-W. et al. (1992) J. Gen. Virol., 73:1131-1141 and  
Lee, C-H et al. (1992) J. Clin. Microbio. 30:1602-1604).  
In addition, at least four more putative genotypes were  
identified thereby providing evidence that the genetic  
5 heterogeneity of HCV was more extensive than previously  
appreciated.

However, while the studies of Bukh et al. (1992a  
and b) provided new and useful information on the genetic  
heterogeneity of HCV, it is widely appreciated by those  
10 skilled in the art that the three structural genes of HCV,  
core (C), envelope (E1) and envelope 2/nonstructural 1  
(E2/NS1) are the most important for the development of  
serologic diagnostics and vaccines since it is the product  
of these genes that constitutes the hepatitis C virion.  
15 Thus, a determination of the nucleotide sequence of one or  
all of the structural genes of a variety of HCV isolates  
would be useful in designing reagents for use in diagnostic  
assays and vaccines since a demonstration of genetic  
heterogeneity in a structural gene(s) of HCV isolates might  
20 suggest that some of the HCV genotypes represent distinct  
serotypes of HCV based upon the previously observed  
relationship between genetic heterogeneity and serologic  
heterogeneity among another group of single-stranded,  
positive-sense RNA viruses, the picornaviruses (Ruechert,  
25 R.R. "Picornaviridae and their replication", in Fields,  
B.N. et al., eds. Virology, New York: Raven Press, Ltd.  
(1990) 507-548).

#### Summary of Invention

30 The present invention relates to cDNAs encoding  
the complete nucleotide sequence of either the envelope 1  
(E1) gene or the core (C) gene of an isolate of human  
hepatitis C virus (HCV).

The present invention also relates to the nucleic  
35 acid and deduced amino acid sequences of these E1 and core

- 6 -

° cDNAs.

It is an object of this invention to provide synthetic nucleic acid sequences capable of directing production of recombinant E1 and core proteins, as well as equivalent natural nucleic acid sequences. Such natural  
5 nucleic acid sequences may be isolated from a cDNA or genomic library from which the gene capable of directing synthesis of the E1 or core proteins may be identified and isolated. For purposes of this application, nucleic acid sequence refers to RNA, DNA, cDNA or any synthetic variant  
10 thereof which encodes for peptides.

The invention also relates to the method of preparing recombinant E1 and core proteins derived from E1 and core cDNA sequences respectively by cloning the nucleic acid encoding either the recombinant E1 or core protein and  
15 inserting the cDNA into an expression vector and expressing the recombinant protein in a host cell.

The invention also relates to isolated and substantially purified recombinant E1 and core proteins and analogs thereof encoded by E1 and core cDNAs respectively.

20 The invention further relates to the use of recombinant E1 and core proteins, either alone, or in combination with each other, as diagnostic agents and as vaccines.

The present invention also relates to the  
25 recombinant production of the core protein of the present invention to contain a second protein on its surface and therefore serve as a carrier in a multivalent vaccine preparation. Further, the present invention relates to the use of the self aggregating core or envelope proteins as a  
30 drug delivery system for anti-virals.

The invention also relates to the use of single-stranded antisense poly- or oligonucleotides derived from E1 or core cDNAs, or from both E1 and core cDNAs, to inhibit expression of hepatitis C E1 and/or core genes.

35 The invention further relates to multiple

- 7 -

- ° computer-generated alignments of the nucleotide and deduced amino acid sequences of the E1 and core cDNAs. These multiple sequence alignments produce consensus sequences which serve to highlight regions of homology and non-homology between sequences found within the same genotype or in different genotypes and hence, these alignments can be used by one skilled in the art to design peptides and oligonucleotides useful as reagents in diagnostic assays and vaccines.

The invention therefore also relates to purified and isolated peptides and analogs thereof derived from E1 and core cDNA sequences.

The invention further relates to the use of these peptides as diagnostic agents and vaccines.

The present invention also encompasses methods of detecting antibodies specific for hepatitis C virus in biological samples. The methods of detecting HCV or antibodies to HCV disclosed in the present invention are useful for diagnosis of infection and disease caused by HCV and for monitoring the progression of such disease. Such methods are also useful for monitoring the efficacy of therapeutic agents during the course of treatment of HCV infection and disease in a mammal.

The invention also provides a kit for the detection of antibodies specific for HCV in a biological sample where said kit contains at least one purified and isolated peptide derived from the E1 or core cDNA sequences. In addition, the invention provides for a kit containing at least one purified and isolated peptide derived from the E1 cDNA sequences and at least one purified and isolated peptide derived from the core cDNA sequences.

The invention further provides isolated and purified genotype-specific oligonucleotides and analogs thereof derived from E1 and core cDNA sequences.

The invention also relates to methods for

- 8 -

- detecting the presence of hepatitis C virus in a mammal, said methods comprising analyzing the RNA of a mammal for the presence of hepatitis C virus. The invention further relates to methods for determining the genotype of hepatitis C virus present in a mammal. This method is useful in determining the proper course of treatment for an HCV-infected patient.

The invention also provides a diagnostic kit for the detection of hepatitis C virus in a biological sample. The kit comprises purified and isolated nucleic acid sequences useful as primers for reverse-transcription polymerase chain reaction (RT-PCR) analysis of RNA for the presence of hepatitis C virus genomic RNA.

The invention further provides a diagnostic kit for the determination of the genotype of a hepatitis C virus present in a mammal. The kit comprises purified and isolated nucleic acid sequences useful as primers for RT-PCR analysis of RNA for the presence of HCV in a biological sample and purified and isolated nucleic acid sequences useful as hybridization probes in determining the genotype of the HCV isolate detected in PCR analysis.

This invention also relates to pharmaceutical compositions useful in prevention or treatment of hepatitis C in a mammal.

#### Description of Figures

Figures 1 A-H show computer generated sequence alignments of the nucleotide sequences of 51 HCV E1 cDNAs. The single letter abbreviations used for the nucleotides shown in Figures 1A-H are those standardly used in the art. Figure 1A shows the alignment of SEQ ID NOs:1-8 to produce a consensus sequence for genotype I/1a. Figure 1B shows the alignment of SEQ ID NOs:9-25 to produce a consensus sequence for genotype II/1b. Figure 1C shows the alignment of SEQ ID NOs:26-29 to produce a consensus sequence for genotype III/2a. Figure 1D shows the alignment of SEQ ID

- 9 -

° NOS:30-33 to produce a consensus sequence for genotype IV/2b. Figure 1E shows the alignment of SEQ ID NOS:35-39 to produce a consensus sequence for genotype V/3a. Figure 1F shows the computer alignment of SEQ ID NOS:42-43 to produce a "consensus" sequence for genotype 4C where the "consensus" sequence given is that of SEQ ID NO:42. Figure 1G shows the alignment of SEQ ID NOS:45-50 to produce a consensus sequence for genotype 5a. The nucleotides shown in capital letters in the consensus sequences of Figures 1A-G are those conserved within a genotype while nucleotides shown in lower case letters in the consensus sequences are those variable within a genotype. In addition, in Figures 1A-E and 1G, when the lower case letter is shown in a consensus sequence, the lower case letter represents the nucleotide found most frequently in the sequences aligned to produce the consensus sequence. In Figure 1F, the lower case letters shown in the consensus sequence are nucleotides in SEQ ID NO:42 which differ from nucleotides found in the same positions in SEQ ID NO:43. Finally, a hyphen at a nucleotide position in the consensus sequences in Figures 1A-G indicates that two nucleotides were found in equal numbers at that position in the aligned sequences. In the aligned sequences, nucleotides are shown in lower case letters if they differed from the nucleotides of both adjacent isolates. Figure 1H shows the alignment of the consensus sequences of Figures 1A-G with SEQ ID NO:34 (genotype 2c), SEQ ID NO:40 (genotype 4a), SEQ ID NO:41 (genotype 4b), SEQ ID NO:44 (genotype 4d) and SEQ ID NO:51 (genotype 6a) to produce a consensus sequence for all twelve genotypes. This consensus sequence is shown as the bottom line of Figure 1H where the nucleotides shown in capital letters are conserved among all genotypes and a blank space indicates that the nucleotide at that position is not conserved among all genotypes.

Figures 2A-H show computer alignments of the deduced amino acid sequences of 51 HCV E1 cDNAs. The

- 10 -

- single letter abbreviations used for the amino acids shown in Figures 2A-H follow the conventional amino acid shorthand for the twenty naturally occurring amino acids. Figure 2A shows the alignment of SEQ ID NOs:52-59 to produce a consensus sequence for genotype I/1a. Figure 2B shows the alignment of SEQ ID NOs:60-76 to produce a consensus sequence for genotype II/1b. Figure 2C shows the alignment of SEQ ID NOs:77-80 to produce a consensus sequence for genotype III/2a. Figure 2D shows the alignment of SEQ ID NOs:81-84 to produce a consensus sequence for genotype IV/2b. Figure 2E shows the alignment of SEQ ID NOs:86-90 to produce a consensus sequence for genotype V/3a. Figure 2F shows the computer alignment of SEQ ID NOs:93-94 to produce a consensus sequence for genotype 4c. Figure 2G shows the alignment of SEQ ID NOs:96-101 to produce a consensus sequence for genotype 5a. The amino acids shown in capital letters in the consensus sequences of Figures 2A-G are those conserved within a genotype while amino acids shown in lower case letters in the consensus sequences are those variable within a genotype. In addition, in Figures 2A-E and 2G when the lower case letter is shown in a consensus sequence, the letter represents the amino acid found most frequently in the sequences aligned to produce the consensus sequence. In Figure 2F, the lower case letters shown in the consensus sequence are amino acids in SEQ ID NO:93 which differ from amino acids found in the same positions in SEQ ID NO:94. Finally, a hyphen at an amino acid position in the consensus sequences of Figures 2A-G indicates that two amino acids were found in equal numbers at that position in the aligned sequences. In the aligned sequences, amino acids are shown in lower case letters if they differed from the amino acids of both adjacent isolates. Figure 2H shows the alignment of the consensus sequences of Figures 2A-G with SEQ ID NO:85 (genotype 2c), SEQ ID NO:91 (genotype 4a), SEQ ID NO:92 (genotype 4b), SEQ ID NO:95 (genotype 4d)



- 11 -

- ° and SEQ ID NO:102 (genotype 6a) to produce a consensus sequence for all twelve genotypes. This consensus sequence is shown as the bottom line of Figure 2H where the amino acids shown in capital letters are conserved among all genotypes and a blank space indicates that the amino acid at that position is not conserved among all genotypes.

Figure 3 shows multiple sequence alignment of the deduced amino acid sequence of the E1 gene of 51 HCV isolates collected worldwide. The consensus sequence of the E1 protein is shown in boldface (top). In the consensus sequence cysteine residues are highlighted with stars, potential N-linked glycosylation sites are underlined, and invariant amino acids are capitalized, whereas variable amino acids are shown in lower case letters. In the alignment, amino acids are shown in lower case letters if they differed from the amino acid of both adjacent isolates. Amino acid residues shown in bold print in the alignment represent residues which at that position in the amino acid sequence are genotype-specific. Amino acids that were invariant among all HCV isolates are shown as hyphens (-) in the alignment. Amino acid positions correspond to those of the HCV prototype sequence (HCV-1, Choo, L. et al. (1991) Proc. Natl. Acad. Sci. USA 88:2451-2455) with the first amino acid of the E1 protein at position 192. The grouping of isolates into 12 genotypes (I/1a, II/1b, III/2a, IV/2b, V/3a, 2c, 4a, 4b, 4c, 4d, 5a and 6a) is indicated.

Figure 4 shows a dendrogram of the genetic relatedness of the twelve genotypes of HCV based on the percent amino acid identity of the E1 gene of the HCV genome. The twelve genotypes shown are designated as I/1a, II/1b, III/2a, IV/2b, V/3a, 2c, 4a, 4b, 4c, 4d, 5a and 6a. The shaded bars represent a range showing the maximum and minimum homology between the amino acid sequence of any one isolate of the genotype indicated and the amino acid sequence of any other isolate.

- 12 -

Figure 5 shows the distribution of the complete E1 gene sequence of 74 HCV isolates into the twelve HCV genotypes in the 12 countries studied. For 51 of these HCV isolates, including 8 isolates of genotype I/1a, 17 isolates of genotype II/1b and 26 isolates comprising the additional 10 genotypes, the complete E1 gene sequence was determined. In the remaining 23 isolates, all of genotypes I/1a and II/1b, the genotype assignment was based on only a partial E1 gene sequence. The partially sequenced isolates did not represent additional genotypes in any of the 12 countries. The number of isolates of a particular genotype is given in each of the 12 countries studied. For ease of viewing, those genotypes designated by two terms (e.g., I/1a) are indicated by the latter term (e.g. 1a). The designations used for each country are: Denmark (DK); Dominican Republic (DR); Germany (D); Hong Kong (HK); India (IND); Sardinia, Italy (S); Peru (P); South Africa (SA); Sweden (SW); Taiwan (T); United States (US); and Zaire (Z). National borders depicted in this figure represent those existing at the time of sampling.

Figures 6A-K show computer generated sequence alignments of the nucleotide sequences of 52 HCV core cDNAs. Single letter abbreviations used for the nucleotides shown in Figures 6A-J are those standardly used in the art. Figure 6A shows the alignment of SEQ ID NOS: 103-108 to produce a consensus sequence for genotype I/1a. Figure 6B shows the alignment of SEQ ID NOS: 109-124 to produce a consensus sequence for genotype II/1b. Figure 6C shows the alignments of the sequences comprising minor genotypes I/1a (SEQ ID NOS: 103-108) and II/1b (SEQ ID NOS: 109-124) to produce a consensus sequence for the major genotype, genotype 1. Figure 6D shows the alignment of SEQ ID NOS: 125-128 to produce a consensus sequence for genotype III/2a. Figure 6E shows the alignment of SEQ ID NOS: 129-133 to produce a consensus sequence for genotype IV/2b. Figure 6F shows the alignment of the sequences of

- 13 -

° minor genotypes III/2a (SEQ ID NOs: 125-128), IV/2b (SEQ ID NOs: 129-133) and 2c (SEQ ID NO: 134) to produce a consensus sequence for the major genotype, genotype 2. Figure 6G shows the alignment of SEQ ID NOs: 135-138 to produce a consensus sequence for genotype V/3a. Figure 6H shows the computer alignment of the sequences of minor genotypes 4a-4f (SEQ ID NOs: 139-145) to produce a consensus sequence for the major genotype, genotype 4. Figure 6I shows the alignment of SEQ ID NOs: 146-153 to produce a consensus sequence for genotype 5a. The nucleotides shown in capital letters in the consensus sequences in Figure 6A-I are those conserved within the genotype while nucleotides shown in lower case letters in the consensus sequences are those variable within a genotype. In addition, when the lower case letter is shown in the consensus sequence, the lower case letter represents the nucleotide found most frequently in the sequences aligned to produce that consensus sequence. Moreover, a hyphen at a nucleotide position in the consensus sequences in Figures 6A-6I indicates that two nucleotides were found in equal numbers at that position in the sequences aligned to produce the consensus sequence. Finally, nucleotides are shown in lower case letters in the sequences aligned to produce each consensus sequence shown in Figures 6A-6I, if they differed from the nucleotides of both adjacent isolates. Figure 6J shows the alignment of the consensus sequences of major genotypes 1 (Figure 6C), 2 (Figure 6F), 3 (Figure 6G), 4 (Figure 6H), 5 (Figure 6I) and 6 (SEQ ID NO: 154) to produce a consensus sequence for all genotypes and Figure 6K shows the alignment of consensus sequences of Figures 6A, 6B, 6D, 6E, 6G and 6I with SEQ ID NO:134 (genotype 2c); SEQ ID NO:139 (genotype 4a); SEQ ID NO:141 (genotype 4b); SEQ ID NO:143 (genotype 4c); SEQ ID NO:145 (genotype 4d); SEQ ID NO:142 (genotype 4e); SEQ ID NO:140 (genotype 4f) and SEQ ID NO:154 (genotype 6a) to produce a consensus sequence for all fourteen genotypes. The

- 14 -

° nucleotides shown in capital letters in the consensus sequences of Figures 6J and 6K are conserved among all genotypes and the nucleotide shown in lower case letter represent the nucleotides found most frequently in the sequences aligned to produce this consensus sequence. In addition, the presence of a hyphen at a nucleotide position in all fourteen sequences aligned in Figure 6K indicates that the nucleotide found at that position in the aligned sequences is the same as nucleotide shown at the corresponding position in the consensus sequences of Figure 6K.

Figures 7A-7J show computer alignments of the deduced amino acid sequences of the 52 HCV core cDNAs. The single letter abbreviations used for the amino acids shown in Figures 7A-7J follow the conventional amino acid short hand for the twenty natural occurring amino acids. Figure 7A shows the alignment of SEQ ID NOS: 155-160 to produce a consensus sequence for genotype I/1a. Figure 7B shows the alignment of SEQ ID NOS: 161-176 to produce a consensus sequence for genotype II/1b. Figure 7C shows the alignment of the sequences comprising minor genotypes I/a (SEQ ID NOS: 155-160) and II/1b (SEQ ID NOS: 161-176) to produce a consensus sequence for the major genotype, genotype 1. Figure 7D shows the alignment of SEQ ID NOS: 177-180 to produce a consensus sequence for genotype III/2a. Figure 7E shows the alignment of SEQ ID NOS: 181-185 to produce a consensus sequence for genotype IV/2b. Figure 7F shows the alignment of the sequences of minor genotypes III/2a (SEQ ID NOS: 177-180), IV/2b (SEQ ID NOS: 181-185) and 2c (SEQ ID NO: 186) to produce a consensus sequence for the major genotype, genotype 2. Figure 7G shows the alignment of SEQ ID NOS: 187-190 to produce a consensus sequence for genotype V/3a. Figure 7H shows the computer alignment of the sequences of minor genotypes 4a-4f (SEQ ID NOS: 191-197) to produce a consensus sequence for the major genotype, genotype 4. Figure 7I shows the alignment of SEQ

- 15 -

• ID NOs: 198-205 to produce a consensus sequence for genotype 5a. The amino acids shown in capital letters in the consensus sequences of Figures 7A-7I are those conserved within the genotype while amino acids shown in lower case letters in the consensus sequences are those variable within the genotype. In addition, when a lower case letter is found in the consensus sequences shown in Figures 7A-7I, the letter represents the amino acid found most frequently in the sequences aligned to produce that consensus sequence. Moreover, a hyphen in an amino acid position in the consensus sequences of Figures 7A-7I indicates that two amino acids were found in equal numbers at that position in the sequences aligned to produce that consensus sequence. Finally, amino acids are shown in lower case letters in the sequences aligned to produce the consensus sequences shown in Figures 7A-7I if these amino acids differed from the amino acids of both adjacent isolates. Figure 7J shows the alignment of the consensus sequences of major genotypes 1 (Figure 7C), 2 (Figure 7F), 3 (Figure 7G), 4 (Figure 7H), 5 (Figure 7I) and 6 (SEQ ID NO: 154) to produce a consensus sequence for all genotypes and Figure 7K shows the alignment of the consensus sequences of Figures 7A, 7B, 7D, 7E, 7G and 7I with SEQ ID NO:186 (genotype 2c), SEQ ID NO:191 (genotype 4a), SEQ ID NO:193 (genotype 4b), SEQ ID NO:195 (genotype 4c), SEQ ID NO:197 (genotype 4d), SEQ ID NO:194 (genotype 4e), SEQ ID NO:192 (genotype 4f) and SEQ ID NO:206 (genotype 6a) to produce a consensus sequence for all fourteen genotypes. The amino acids shown in capital letters in the consensus sequences shown in Figures 7J and 7K are conserved among all genotypes while the amino acids shown in lower case letters represent amino acids found most frequently in the sequences aligned to produce this consensus sequence. In addition, the presence of a hyphen at an amino acid position in all fourteen sequences aligned in Figure 7K indicates that the amino acid found at that position in the

- 16 -

- ° aligned sequences is the same as the amino acid shown at the corresponding position in the consensus sequence of Figure 7K.

Figure 8 shows phylogenetic trees illustrating the calculated evolutionary relationships of the different HCV isolates based upon the C gene sequence of 52 HCV isolates and the E1 gene sequence of 51 HCV isolates, respectively. The phylogenetic trees were constructed by the unweighted pair-group method with arithmetic mean (Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia University Press, New York, N.Y.), pp 287-326) using the computer software package "Gene Works" from IntelliGenetics. The lengths of the horizontal lines connecting the sequences, given in absolute values from 0 to 1, are proportional to the estimated genetic distances between the sequences. Genotype designations of HCV isolates are indicated. In 45 HCV isolates, both the C and the E1 gene sequences were determined.

#### Detailed Description Of Invention

The present invention relates to cDNAs encoding the complete nucleotide sequence of the envelope 1 (E1) and core genes of isolates of human hepatitis C virus (HCV). The E1 cDNAs of the present invention were obtained as follows. Viral RNA was extracted from serum collected from humans infected with hepatitis C virus and the viral RNA was then reverse transcribed and amplified by polymerase chain reaction using primers deduced from the sequence of the HCV strain H-77 (Ogata, N. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:3392-3396). The amplified cDNA was then isolated by gel electrophoresis and sequenced.

The present invention further relates to the nucleotide sequences of the cDNAs encoding the E1 gene of 51 HCV isolates. These nucleotide sequences are shown in the sequence listing as SEQ ID NO:1 through SEQ ID NO:51.

The abbreviations used for the nucleotides are

- 17 -

- ° those standardly used in the art.

The deduced amino acid sequence of each of SEQ ID NO:1 through SEQ ID NO:51 are presented in the sequence listing as SEQ ID NO:52 through SEQ ID NO:102 where the amino acid sequence in SEQ ID NO:52 is deduced from the nucleotide sequence shown in SEQ ID NO:1, the amino acid sequence shown in SEQ ID NO:53 is deduced from the nucleotide sequence shown in SEQ ID NO:2 and so on. The deduced amino acid sequence of each of SEQ ID Nos:52-102 starts at nucleotide 1 of the corresponding nucleic acid sequence shown in SEQ ID NOS:1-51 and extends 575 nucleotides to a total length of 576 nucleotides.

The three letter abbreviations used in SEQ ID Nos:52-102 follow the conventional amino acid shorthand for the twenty naturally occurring amino acids.

The present invention also relates to the nucleotide sequences of the cDNAs encoding the core gene of 52 HCV isolates. These nucleotide sequences are shown in the sequence listing as SEQ ID NO:103 through SEQ ID NO:154.

The core cDNAs of the present invention were obtained as follows. Viral RNA was extracted from serum and reversed transcribed as described above for cloning of the E1 cDNAs. The core cDNAs of the present invention were then amplified by polymerase chain reaction using primers deduced from previously determined sequences that flank the core gene (Bukh et al. (1992)) Proc. Natl. Acad. Sci. U.S.A., 89: 4942-4946; Bukh et al. (1993) Proc. Natl. Acad. Sci. U.S.A., 90: 8234-8238).

The deduced amino acid sequence of each of SEQ ID NO:103 through SEQ ID NO:154 are presented in the sequence listing as SEQ ID NO:155 through SEQ ID NO:206 where the amino acid sequence in SEQ ID NO:155 is deduced from the nucleotide sequence shown in SEQ ID NO:103, the amino acid sequence shown in SEQ ID NO:156 is deduced from the nucleotide sequence shown in SEQ ID NO:104 and so on. The

- 18 -

- ° deduced amino acid sequence of each of SEQ ID NOs: 155-206 starts at nucleotide 1 of the corresponding nucleotide sequence shown in SEQ ID NOs:103-154 and extends 572 nucleotides to a total length of 573 nucleotides.

Preferably, the E1 and core proteins and peptides of the present invention are substantially homologous to, and most preferably biologically equivalent to, native HCV E1 and core proteins and peptides. By "biologically equivalent" as used throughout the specification and claims, it is meant that the compositions are immunogenically equivalent to the native E1 and core proteins and peptides. The E1 and core proteins and peptides of the present invention may also stimulate the production of protective antibodies upon injection into a mammal that would serve to protect the mammal upon challenge with HCV. By "substantially homologous" as used throughout the ensuing specification and claims to describe E1 and core proteins and peptides, it is meant a degree of homology in the amino acid sequence of the E1 and core proteins and peptides to the native E1 and core proteins and peptides respectively. Preferably the degree of homology is in excess of 90, preferably in excess of 95, with a particularly preferred group of proteins being in excess of 99 homologous with the native E1 or core proteins and peptides.

Variations are contemplated in the cDNA sequences shown in SEQ ID NO:1 through SEQ ID NO:51 and in SEQ ID NO:103 through SEQ ID NO:154 which will result in a nucleic acid sequence that is capable of directing production of analogs of the corresponding protein shown in SEQ ID NO:52 through SEQ ID NO:102 and in SEQ ID NO:155 through SEQ ID NO:206. It should be noted that the cDNA sequences set forth above represent a preferred embodiment of the present invention. Due to the degeneracy of the genetic code, it is to be understood that numerous choices of nucleotides may be made that will lead to a DNA sequence capable of



- 19 -

- ° directing production of the instant protein or its analogs. As such, DNA sequences which are functionally equivalent to the sequence set forth above or which are functionally equivalent to sequences that would direct production of analogs of the E1 and core proteins produced pursuant to the amino acid sequences set forth above, are intended to be encompassed within the present invention.

The term analog as used throughout the specification or claims to describe the E1 and core proteins and peptides of the present invention, includes any protein or peptide having an amino acid residue sequence substantially identical to a sequence specifically shown herein in which one or more residues have been conservatively substituted with a biologically equivalent residue. Examples of conservative substitutions include the substitution of one polar (hydrophobic) residue such as isoleucine, valine, leucine or methionine for another, the substitution of one polar (hydrophilic) residue for another such as between arginine and lysine, between glutamine and asparagine, between glycine and serine, the substitution of one basic residue such as lysine, arginine or histidine for another, or the substitution of one acidic residue, such as aspartic acid or glutamic acid for another.

The phrase "conservative substitution" also includes the use of a chemically derivatized residue in place of a non-derivatized residue provided that the resulting protein or peptide is biologically equivalent to the native E1 or core protein or peptide.

"Chemical derivative" refers to an E1 or core protein or peptide having one or more residues chemically derivatized by reaction of a functional side group. Examples of such derivatized molecules, include but are not limited to, those molecules in which free amino groups have been derivatized to form amine hydrochlorides, p-toluene sulfonyl groups, carbobenzoxy groups, t-butyloxycarbonyl groups, chloracetyl groups or formyl groups. Free carboxyl

- 20 -

groups may be derivatized to form salts, methyl and ethyl esters or other types of esters or hydrazides. Free hydroxyl groups may be derivatized to form O-acyl or O-alkyl derivatives. The imidazole nitrogen of histidine may be derivatized to form N-imbenzylhistidine. Also included as chemical derivatives are those proteins or peptides which contain one or more naturally-occurring amino acid derivatives of the twenty standard amino acids. For examples: 4-hydroxyproline may be substituted for proline; 5-hydroxylysine may be substituted for lysine; 3-methylhistidine may be substituted for histidine; homoserine may be substituted for serine; and ornithine may be substituted for lysine. The E1 and core proteins and peptide of the present invention also includes any protein or peptide having one or more additions and/or deletions of residues relative to the sequence of a peptide whose sequence is shown herein, so long as the peptide is biologically equivalent to the native E1 or core protein or peptide.

The present invention also includes a recombinant DNA method for the manufacture of HCV E1 and core proteins. In this method, natural or synthetic nucleic acid sequences may be used to direct the production of E1 and core proteins.

In one embodiment of the invention, the method comprises:

- (a) preparation of a nucleic acid sequence capable of directing a host organism to produce HCV E1 or core protein;
- (b) cloning the nucleic acid sequence into a vector capable of being transferred into and replicated in a host organism, such vector containing operational elements for the nucleic acid sequence;
- (c) transferring the vector containing the nucleic acid and operational elements into a host organism capable of expressing the protein;

- 21 -

° (d) culturing the host organism under conditions appropriate for amplification of the vector and expression of the protein; and

(e) harvesting the protein.

5 In another embodiment of the invention, the method for the recombinant DNA synthesis of an HCV E1 protein encoded by any one of the nucleic acid sequences shown in SEQ ID NOS:1-51 comprises:

10 (a) culturing a transformed or transfected host organism containing a nucleic acid sequence capable of directing the host organism to produce a protein, under conditions such that the protein is produced, said protein exhibiting substantial homology to a native E1 protein isolated from HCV having the amino acid sequence according to any one of the amino acid sequences shown in SEQ ID  
15 NOS:52-102 or combinations thereof.

In one embodiment, the RNA sequence of an HCV isolate was isolated and converted to cDNA as follows. Viral RNA is extracted from a biological sample collected from human subjects infected with hepatitis C and the viral  
20 RNA is then reverse transcribed and amplified by polymerase chain reaction using primers deduced from the sequence of HCV strain H-77 (Ogata et al. (1991)). Preferred primer sequences are shown as SEQ ID NOS:207-212 in the sequence listing. Once amplified, the PCR fragments are isolated by  
25 gel electrophoresis and sequenced.

In an alternative embodiment, the above method may be utilized for the recombinant DNA synthesis of an HCV core protein encoded by any one of the nucleic acid sequences shown in SEQ ID NOS: 103-154, where the protein  
30 produced by this method exhibits substantial homology to a native core protein isolated from HCV having amino acid sequence according to any one of the amino acid sequences shown in SEQ ID NOS: 155-206 or combinations thereof.

35 The vectors contemplated for use in the present invention include any vectors into which a nucleic acid

- 22 -

° sequence as described above can be inserted, along with any preferred or required operational elements, and which vector can then be subsequently transferred into a host organism and replicated in such organisms. Preferred vectors are those whose restriction sites have been well documented and which contain the operational elements preferred or required for transcription of the nucleic acid sequence.

The "operational elements" as discussed herein include at least one promoter, at least one operator, at least one leader sequence, at least one terminator codon, and any other DNA sequences necessary or preferred for appropriate transcription and subsequent translation of the vector nucleic acid. In particular, it is contemplated that such vectors will contain at least one origin of replication recognized by the host organism along with at least one selectable marker and at least one promoter sequence capable of initiating transcription of the nucleic acid sequence.

In construction of the recombinant expression vectors of the present invention, it should additionally be noted that multiple copies of the nucleic acid sequence of interest (either E1 or core) and its attendant operational elements may be inserted into each vector. In such an embodiment, the host organism would produce greater amounts per vector of the desired E1 or core protein. The number of multiple copies of the nucleic acid sequence which may be inserted into the vector is limited only by the ability of the resultant vector due to its size, to be transferred into and replicated and transcribed in an appropriate host microorganism.

Of course, those skilled in the art would readily understand that copies of both core and E1 nucleic acid sequence may be inserted into single vector such that a host organism transformed or transfected with said vector would produce both the desired E1 and core proteins. For

- 23 -

° example, a polysistronic vector in which multiple different E1 and/or core proteins may be expressed from a single vector is created by placing expression of each protein under control of an internal ribosomal entry site (IRES) (Molla, A. et al. Nature, 356:255-257 (1992); Gong, S.K. et al. J. of Virol., 263:1651-1660 (1989)).

5 In another embodiment, restriction digest fragments containing a coding sequence for E1 or core proteins can be inserted into a suitable expression vector that functions in prokaryotic or eukaryotic cells. By  
10 suitable is meant that the vector is capable of carrying and expressing a complete nucleic acid sequence coding for an E1 or core protein. Preferred expression vectors are those that function in a eukaryotic cell. Examples of such vectors include but are not limited to vaccinia virus  
15 vectors, adenovirus or herpes viruses. A preferred vector is the baculovirus transfer vector, pBlueBac.

In yet another embodiment, the selected recombinant expression vector may then be transfected into a suitable eukaryotic cell system for purposes of  
20 expressing the recombinant protein. Such eukaryotic cell systems include but are not limited to cell lines such as HeLa, MRC-5 or CV-1. A preferred eukaryotic cell system is SF9 insect cells.

25 The expressed recombinant protein may be detected by methods known in the art including, but not limited to, Coomassie blue staining and Western blotting.

The present invention also relates to substantially purified and isolated recombinant E1 and core proteins. In one embodiment, the recombinant protein  
30 expressed by the SF9 cells can be obtained as a crude lysate or it can be purified by standard protein purification procedures known in the art which may include differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel  
35 electrophoresis and affinity and immunoaffinity

- 24 -

° chromatography. The recombinant protein may be purified by passage through a column containing a resin which has bound thereto antibodies specific for the open reading frame (ORF) protein.

5 The present invention further relates to the use of recombinant E1 and core proteins as diagnostic agents and vaccines. In one embodiment, the expressed recombinant proteins of this invention can be used in immunoassays for diagnosing or prognosing hepatitis C in a mammal. For the purposes of the present invention, "mammal" as used  
10 throughout the specification and claims, includes, but is not limited to humans, chimpanzees, other primates and the like. In a preferred embodiment, the immunoassay is useful in diagnosing hepatitis C infection in humans.

Immunoassays of the present invention may be  
15 those commonly used by those skilled in the art including, but not limited to, radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme immunoassay, chemiluminescent assay, immunohistochemical assay, immunoprecipitation and the like. Standard techniques  
20 known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Edition, Rose and Bigazzi, eds., John Wiley and Sons, 1980 and Campbell et al., Methods of Immunology, W.A. Benjamin, Inc., 1964, both of which are incorporated herein by reference. Such assays may be a  
25 direct, indirect, competitive, or noncompetitive immunoassay as described in the art (Oellerich, M. 1984. J. Clin. Chem. Clin. BioChem 22:895-904) Biological samples appropriate for such detection assays include, but are not limited to serum, liver, saliva, lymphocytes or other  
30 mononuclear cells.

In a preferred embodiment, test serum is reacted with a solid phase reagent having surface-bound recombinant HCV E1 and/or core protein(s) as antigen(s). The solid surface reagent can be prepared by known techniques for  
35 attaching protein to solid support material. These

- 25 -

attachment methods include non-specific adsorption of the protein to the support or covalent attachment of the protein to a reactive group on the support. After reaction of the antigen with anti-HCV antibody, unbound serum components are removed by washing and the antigen-antibody complex is reacted with a secondary antibody such as labelled anti-human antibody. The label may be an enzyme which is detected by incubating the solid support in the presence of a suitable fluorimetric or calorimetric reagent. Other detectable labels may also be used, such as radiolabels or colloidal gold, and the like.

The HCV E1 and/or core proteins and analogs thereof may be prepared in the form of a kit, alone, or in combinations with other reagents such as secondary antibodies, for use in immunoassays.

In yet another embodiment the recombinant E1 and core proteins or analogs thereof can be used as a vaccine to protect mammals against challenge with hepatitis C. The vaccine, which acts as an immunogen, may be a cell, cell lysate from cells transfected with a recombinant expression vector or a culture supernatant containing the expressed protein. Alternatively, the immunogen is a partially or substantially purified recombinant protein. In yet another embodiment, the immunogen may be a fusion protein comprising core protein and a second, non-core protein joined together such that the core portion of the fusion protein will aggregate and "trap" the second protein on the surface of the particle produced by aggregation of the core protein. (Molecular Biology of the Hepatitis B Virus", McLachlan, A. (1991) CRC Press, Boca Raton, Fla.). Alternatively, the core protein could be mixed with the second protein in vitro to produce particles in which all or part of the second protein was exposed on the surface of the particle. Such particles would then serve as a carrier in a multi-valent vaccine preparation. Second proteins or parts thereof which could be mixed with or fused to the

- 26 -

- core protein include, but are not limited to, HCV E1 and hepatitis B surface antigen.

While it is possible for the immunogen to be administered in a pure or substantially pure form, it is preferable to present it as a pharmaceutical composition, formulation or preparation.

The formulations of the present invention, both for veterinary and for human use, comprise an immunogen as described above, together with one or more pharmaceutically acceptable carriers and optionally other therapeutic ingredients. The carrier(s) must be "acceptable" in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. The formulations may conveniently be presented in unit dosage form and may be prepared by any method well-known in the pharmaceutical art.

All methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In general, the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product into the desired formulation.

Formulations suitable for intravenous intramuscular, subcutaneous, or intraperitoneal administration conveniently comprise sterile aqueous solutions of the active ingredient with solutions which are preferably isotonic with the blood of the recipient. Such formulations may be conveniently prepared by dissolving the solid active ingredient in water containing physiologically compatible substances such as sodium chloride (e.g. 0.1-2.0m), glycine, and the like, and having a buffered pH compatible with physiological conditions to produce an aqueous solution, and rendering said solution sterile. These may be present in unit or multi-dose containers, for



- 27 -

- ° example, sealed ampules or vials.

Th formulations of the present invention may incorporate a stabilizer. Illustrative stabilizers are preferably incorporated in an amount of 0.10-10,000 parts by weight per part by weight of immunogens. If two or more  
5 stabilizers are to be used, their total amount is preferably within the range specified above. These stabilizers are used in aqueous solutions at the appropriate concentration and pH. The specific osmotic pressure of such aqueous solutions is generally in the  
10 range of 0.1-3.0 osmoles, preferably in the range of 0.8-1.2. The pH of the aqueous solution is adjusted to be within the range of 5.0-9.0, preferably within the range of 6-8. In formulating the immunogen of the present invention, an anti-adsorption agent may be used.

15 Additional pharmaceutical methods may be employed to control the duration of action. Controlled release preparations may be achieved through the use of polymer to complex or adsorb the proteins or their derivatives. The controlled delivery may be exercised by selecting  
20 appropriate macromolecules (for example polyester, polyamino acids, polyvinyl pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine sulfate) and the concentration of macromolecules as well as the methods of  
25 incorporation in order to control release. Another possible method to control the duration of action by controlled-release preparations is to incorporate the proteins, protein analogs or their functional derivatives, into particles of a polymeric material such as polyesters,  
30 polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by  
35 interfacial polymerization, for example,

- 28 -

- ° hydroxymethylcellulose or gelatin-microcapsules and poly (methylemethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions.

5                   When oral preparations are desired, the compositions may be combined with typical carriers, such as lactose, sucrose, starch, talc, magnesium stearate, crystalline cellulose, methyl cellulose, carboxymethyl cellulose, glycerin, sodium alginate or gum arabic among  
10 others.

                  The E1 and core proteins of the present invention may also be used as a delivery system for anti-virals to prevent or attenuate HCV infection in a mammal by utilizing the property of both proteins to self-aggregate in vitro to  
15 "trap" the antiviral within the particles produced via aggregation of the core and E1 proteins. Examples of anti-virals which could be delivered by such a system include, but are not limited to antisense DNA or RNAs.

                  Vaccination can be conducted by conventional  
20 methods. For example, the immunogen or immunogens (e.g. the E1 protein may be administered alone or in combination with the E1 proteins derived from other isolates of HCV) can be used in a suitable diluent such as saline or water, or complete or incomplete adjuvants. Further, the  
25 immunogen(s) may or may not be bound to a carrier to make the protein(s) immunogenic. Examples of such carrier molecules include but are not limited to bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), tetanus toxoid, and the like. The immunogen(s) can be administered  
30 by any route appropriate for antibody production such as intravenous, intraperitoneal, intramuscular, subcutaneous, and the like. The immunogen(s) may be administered once or at periodic intervals until a significant titer of anti-HCV antibody is produced. The antibody may be detected in the  
35 serum using an immunoassay.

- 29 -

° In yet another embodiment, the immunogen may be nucleic acid sequence capable of directing host organism synthesis of E1 and/or core protein(s). Such nucleic acid sequence may be inserted into a suitable expression vector by methods known to those skilled in the art. Expression  
5 vectors suitable for producing high efficiency gene transfer in vivo include retroviral, adenoviral and vaccinia viral vectors. Operational elements of such expression vectors are disclosed previously in the present specification and are known to one skilled in the art.  
10 Such expression vectors can be administered intravenously, intramuscularly, subcutaneously, intraperitoneally or orally.

In an alternative embodiment, direct gene transfer may be accomplished via intramuscular injection  
15 of, for example, plasmid-based eukaryotic expression vectors containing a nucleic acid sequence capable of directing host organism synthesis of E1 and/or core protein(s). Such an approach has previously been utilized to produce the hepatitis B surface antigen in vivo and  
20 resulted in an antibody response to the surface antigen (Davis, H.L. et al. (1993) Human molecular Genetics, 2:1847-1851; see also Davis et al. (1993) Human Gene Therapy, 4:151-159 and 733-740).

Doses of E1 and/or core protein(s)-encoding  
25 nucleic acid sequence effective to elicit a protective antibody response against HCV infection range from about 1 to about 500 µg. A more preferred range being about 1 to about 500 µg.

The E1 and/or core proteins and expression  
30 vectors containing a nucleic acid sequence capable of directing host organism synthesis of E1 and/or core protein(s) may be supplied in the form of a kit, alone, or in the form of a pharmaceutical composition as described above.

35 The administration of the immunogen(s) of the

- 30 -

° present invention may be for either a prophylactic or therapeutic purpose. When provided prophylactically, the immunogen(s) is provided in advance of any exposure to HCV or in advance of any symptom of any symptoms due to HCV infection. The prophylactic administration of the immunogen serves to prevent or attenuate any subsequent infection of HCV in a mammal. When provided therapeutically, the immunogen(s) is provided at (or shortly after) the onset of the infection or at the onset of any symptom of infection or disease caused by HCV. The therapeutic administration of the immunogen(s) serves to attenuate the infection or disease.

In addition to use as a vaccine, the compositions can be used to prepare antibodies to HCV E1 and core proteins. The antibodies can be used directly as antiviral agents or they may be used in immunoassays disclosed herein to detect HCV E1 and core proteins present in patient sera.. To prepare antibodies, a host animal is immunized using the E1 and/or core proteins native to the virus particle bound to a carrier as described above for vaccines. The host serum or plasma is collected following an appropriate time interval to provide a composition comprising antibodies reactive with the E1 or core protein of the virus particle. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE Sephadex, or other techniques known to those skilled in the art. The antibodies are substantially free of many of the adverse side effects which may be associated with other anti-viral agents such as drugs.

The antibody compositions can be made even more compatible with the host system by minimizing potential adverse immune system responses. This is accomplished by removing all or a portion of the Fc portion of a foreign species antibody or using an antibody of the same species as the host animal, for example, the use of antibodies from

- 31 -

human/human hybridomas. Humanized antibodies (i.e., nonimmunogenic in a human) may be produced, for example, by replacing an immunogenic portion of an antibody with a corresponding, but nonimmunogenic portion (i.e., chimeric antibodies). Such chimeric antibodies may contain the reactive or antigen-binding portion of an antibody from one species and the Fc portion of an antibody (nonimmunogenic) from a different species. Examples of chimeric antibodies, include but are not limited to, non-human mammal-human chimeras, rodent-human chimeras, murine-human and rat-human chimeras (Robinson et al., International Patent Application 184,187; Taniguchi M., European Patent Application 171,496; Morrison et al., European Patent Application 173,494; Neuberger et al., PCT Application WO 86/01533; Cabilly et al., 1987 Proc. Natl. Acad. Sci. USA 84:3439; Nishimura et al., 1987 Canc. Res. 47:999; Wood et al., 1985 Nature 314:446; Shaw et al., 1988 J. Natl. Cancer Inst. 80:15553, all incorporated herein by reference).

General reviews of "humanized" chimeric antibodies are provided by Morrison S., 1985 Science 229:1202 and by Oi et al., 1986 BioTechniques 4:214.

Suitable "humanized" antibodies can be alternatively produced by CDR or CEA substitution (Jones et al., 1986 Nature 321:552; Verhoeyan et al., 1988 Science 239:1534; Biedler et al. 1988 J. Immunol. 141:4053, all incorporated herein by reference).

The antibodies or antigen binding fragments may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of the PCT patent applications; publication number WO 901443, WO901443, and WO 9014424 and in Huse et al., 1989 Science 246:1275-1281.

The antibodies can also be used as a means of enhancing the immune response. The antibodies can be administered in amount similar to those used for other therapeutic administrations of antibody. For example,

- 32 -

- ° normal immune globulin is administered at 0.02-0.1 ml/lb body weight during the early incubation period of other viral diseases such as rabies, measles, and hepatitis B to interfere with viral entry into cells. Thus, antibodies reactive with the HCV E1 and/or core proteins can be passively administered alone or in conjunction with another anti-viral agent to a host infected with an HCV to enhance the immune response and/or the effectiveness of an antiviral drug.

Alternatively, anti-HCV E1 antibodies and anti-HCV core antibodies can be induced by administered anti-idiotypic antibodies as immunogens. Conveniently, a purified anti-HCV E1 or anti-HCV core antibody preparation prepared as described above is used to induce anti-idiotypic antibody in a host animal, the composition is administered to the host animal in a suitable diluent. Following administration, usually repeated administration, the host produces anti-idiotypic antibody. To eliminate an immunogenic response to the Fc region, antibodies produced by the same species as the host animal can be used or the Fc region of the administered antibodies can be removed. Following induction of anti-idiotypic antibody in the host animal, serum or plasma is removed to provide an antibody composition. The composition can be purified as described above for anti-HCV E1 and anti-HCV core antibodies, or by affinity chromatography using anti-HCV E1 or anti-HCV core antibodies bound to the affinity matrix. The anti-idiotypic antibodies produced are similar in conformation to the authentic HCV E1 or core protein and may be used to prepare an HCV vaccine rather than using an HCV E1 or core protein.

When used as a means of inducing anti-HCV virus antibodies in an animal, the manner of injecting the antibody is the same as for vaccination purposes, namely intramuscularly, intraperitoneally, subcutaneously or the like in an effective concentration in a physiologically suitable diluent with or without adjuvant. One or more

- 33 -

- booster injections may be desirable.

The HCV E1 and core proteins of the invention are also intended for use in producing antiserum designed for pre- or post-exposure prophylaxis. Here an E1 or core protein, or mixture of E1 and/or core proteins is  
5 formulated with a suitable adjuvant and administered by injection to human volunteers, according to known methods for producing human antisera. Antibody response to the injected proteins is monitored, during a several-week period following immunization, by periodic serum sampling  
10 to detect the presence of anti-HCV E1 and/or anti-HCV core serum antibodies, using an immunoassay as described herein.

The antiserum from immunized individuals may be administered as a pre-exposure prophylactic measure for individuals who are at risk of contracting infection. The  
15 antiserum is also useful in treating an individual post-exposure, analogous to the use of high titer antiserum against hepatitis B virus for post-exposure prophylaxis.

For both in vivo use of antibodies to HCV virus-like particles and proteins and anti-idiotypic antibodies and diagnostic use, it may be preferable to use monoclonal  
20 antibodies. Monoclonal anti-HCV E1 and anti-HCV core protein antibodies or anti-idiotypic antibodies can be produced as follows. The spleen or lymphocytes from an immunized animal are removed and immortalized or used to  
25 prepare hybridomas by methods known to those skilled in the art. (Goding, J.W. 1983. Monoclonal Antibodies: Principles and Practice, Pladermic Press, Inc., NY, NY, pp. 56-97). To produce a human-human hybridoma, a human  
30 lymphocyte donor is selected. A donor known to be infected with HCV (where infection has been shown for example by the presence of anti-virus antibodies in the blood or by virus culture) may serve as a suitable lymphocyte donor.  
Lymphocytes can be isolated from a peripheral blood sample or spleen cells may be used if the donor is subject to  
35 splenectomy. Epstein-Barr virus (EBV) can be used to

- 34 -

- ° immortalize human lymphocytes or a human fusion partner can be used to produce human-human hybridomas. Primary in vitro immunization with peptides can also be used in the generation of human monoclonal antibodies.

Antibodies secreted by the immortalized cells are  
5 screened to determine the clones that secrete antibodies of the desired specificity. For monoclonal anti-E1 and anti-core antibodies, the antibodies must bind to HCV E1 and core proteins respectively. For monoclonal anti-idiotypic antibodies, the antibodies must bind to anti-E1 and anti-core protein antibodies respectively. Cells producing  
10 antibodies of the desired specificity are selected.

The present invention also relates to the use of single-stranded antisense poly- or oligonucleotides derived from nucleotide sequences substantially homologous to those  
15 shown in SEQ ID NOs:1-51 to inhibit the expression of hepatitis C E1 genes. The present invention further relates to the use of single-stranded anti-sense poly- or oligo-nucleotides derived from nucleotide sequences substantially homologous to those shown in SEQ ID NOs:103-  
20 154 to inhibit the expression of hepatitis C core genes. Alternatively, the anti-sense poly- or oligo-nucleotides may be complementary to both the E1 and core genes and hence, inhibit the expression of both hepatitis C E1 and core genes. By substantially homologous as used throughout  
25 the specification and claims to describe the nucleic acid sequences of the present invention, is meant a level of homology between the nucleic acid sequence and the SEQ ID NOs. referred to in the above sentence. Preferably, the level of homology is in excess of 80%, more preferably in excess of 90%, with a preferred nucleic acid sequence being  
30 in excess of 95% homologous with the DNA sequence shown in the indicated SEQ ID NO. These anti-sense poly- or oligonucleotides can be either DNA or RNA. The targeted sequence is typically messenger RNA and more preferably, a  
35 single sequence required for processing or translation of



- 35 -

- ° the RNA. The anti-sense poly- or oligonucleotides can be conjugated to a polycation such as polylysine as disclosed in Lemaitre, M. et al. ((1989) Proc. Natl. Acad. Sci. USA 84:648-652) and this conjugate can be administered to a mammal in an amount sufficient to hybridize to and inhibit the function of the messenger RNA.

The present invention further relates to multiple computer-generated alignments of the nucleotide and deduced amino acid sequences shown in SEQ ID NOS:1-206. Computer analysis of the nucleotide sequences shown in SEQ ID NOS:1-51 and 103-154 and of the deduced amino acid sequences shown in SEQ ID NOS:52-102 and 155-206 can be carried out using commercially available computer programs known to one skilled in the art.

In one embodiment, computer analysis of SEQ ID NOS:1-51 by the program GENALIGN (Intelligenetics, Inc. Mountainview, CA) results in distribution of the 51 HCV E1 sequences into twelve genotypes based upon the degree of variation of the sequences. For the purposes of the present invention, the nucleotide sequence identity of E1 cDNAs of HCV isolates of the same genotype is in the range of about 85% to about 100% whereas the identity of E1 cDNA sequences of different genotypes is in the range of about 50% to about 80%.

The grouping of SEQ ID NOS:1-51 into twelve HCV genotypes is shown below.

	<u>SEQ ID NOS:</u>	<u>Genotypes</u>
	1-8	I/1a
	9-25	II/1b
	26-29	III/2a
	30-33	IV/2b
30	34	2c
	35-39	V/3a
	40	4a
	41	4b
	42-43	4c
	44	4d
	45-50	5a
35	51	6a

- 36 -

° For those genotypes containing more than one E1 nucleotide sequence, computer alignment of the constituent nucleotide sequences of the genotype was conducted using GENALIGN in order to produce a consensus sequence for each genotype. These alignments and their resultant consensus sequences are shown in Figures 1A-G for the seven genotypes (I/1a, II/1b, III/2a, IV/2b, V/3a, 4c and 5a) which comprise more than one nucleotide sequence. Further alignment of the consensus sequences of Figures 1A-G with SEQ ID NO:34 (genotype 2c), SEQ ID NO:40 (genotype 4a), SEQ ID NO:41 (genotype 4b), SEQ ID NO:44 (genotype 4d) and SEQ ID NO:51 (genotype 6a) produces a consensus sequence for all twelve genotypes as shown in Figure 1H. The multiple alignments of nucleotide sequences shown in Figures 1A-H produce consensus sequences which serve to highlight regions of homology and non-homology between sequences found within the same genotype or in different genotypes and hence, these alignments can be used by one skilled in the art to design oligonucleotides useful as reagents in diagnostic assays for HCV.

20 Examples of purified and isolated oligonucleotide sequences derived from the consensus sequences shown in Figures 1A-H include, but are not limited to, SEQ ID NOs:213-239 where these oligonucleotides are useful as "genotype-specific" primers and probes since these oligonucleotides can hybridize specifically to the nucleotide sequence of the E1 gene of HCV isolates belonging to a single genotype. The genotype-specificity of the oligonucleotides shown in SEQ ID NOs:213-239 is as follows: SEQ ID NOs:213-214 are specific for genotype I/1a; SEQ ID NOs:215-216 are specific for genotype II/1b; SEQ ID NOs:217-218 are specific for genotype III/2a; SEQ ID NOs:219-220 are specific for genotype IV/2b; SEQ ID NOs:221-223 are specific for genotype 2c; SEQ ID NOs:224-226 are specific for genotype V/3a; SEQ ID NOs:227-228 are specific for genotype 4a; SEQ ID NOs:229-230 are specific

- 37 -

- for genotype 4b; SEQ ID NOS:231-232 are specific for genotype 4c; SEQ ID NOS:233-234 are specific for genotype 4d; SEQ ID NOS:235-236 are specific for genotype 5a and SEQ ID NOS:237-239 are specific for genotype 6a.

In another embodiment, the computer analysis of SEQ ID NOS:103-154 by the program GENALIGN results in distribution of the 52 HCV core sequences into 14 genotypes based upon the degree of variation of the sequences.

The grouping of SEQ ID NOS:103-154 into 14 HCV genotypes is shown below.

	<u>SEQ ID NOS:</u>	<u>Genotypes</u>
	103-108	I/1a
	109-124	II/1b
	125-128	III/2a
	129-133	IV/2b
	134	2c
15	135-138	V/3a
	139	4a
	141	4b
	143	4c
	144	4c
	145	4d
	142	4e
	140	4f
20	146-153	5a
	154	6a

These 14 genotypes can be further grouped into 6 major genotypes designated genotypes 1-6 where genotype 1 comprises the sequences contained in minor genotypes I/1a and II/1b; genotype 2 comprises the sequences contained in minor genotypes III/2a, IV/2b and 2c; genotype 3 comprises sequences contained in genotype V/3a; genotype 4 comprises sequences contained in minor genotypes 4a-4f; genotype 5 comprises the sequences contained in genotype 5a and genotype 6 comprises the sequence contained in genotype 6a. Computer alignment of the constituent nucleotide sequences of the core cDNAs falling within genotypes I/1a, II/1b, III/2a, IV/2b, V/3a and 5a, to produce a consensus sequence for each of these genotypes is shown in Figures 6A (I/1a),

- 38 -

° 6B (II/1b), 6D (III/2a), 6E (IV/2b), 6G (V/3a) and 6I (5a). The alignment of the sequences found in minor genotypes I/1a and II/1b to produce a consensus sequence for major genotype 1 is shown in Figure 6C. The alignment of the sequences contained in minor genotypes III/2a, IV/2b and 2c to produce a consensus sequence for major genotype 2 is shown in Figure 6F. The alignment of the nucleotide sequences contained in minor genotypes 4a-4f to produce a consensus sequence for major genotype 4 is shown in Figure 6H. Further alignment of the consensus sequences shown in Figures 6C, 6F, 6G, 6H and 6I with SEQ ID NO:154 (genotype 6a/major genotype 6) to produce a consensus sequence for all genotypes is shown in Figure 6J and alignment of the consensus sequences shown in Figures 6A, 6B, 6D, 6E, 6G and 6I with 4a), SEQ ID NO:141 (genotype 4b), SEQ ID NO:143 (genotype 4c), SEQ ID NO:145 (genotype 4d), SEQ ID NO:142 (genotype 4e), SEQ ID NO:140 (genotype 4f) and SEQ ID NO:154 (genotype 6a) to produce a consensus sequence for all fourteen genotypes is shown in Figure 6K. As with the alignments of the envelope (E1) nucleotide sequences, the consensus sequences shown in Figures 6A-6K serve to highlight regions of homology and non-homology between sequences found within the same genotype or in different genotypes and hence, can be used by one skilled in the art to design oligonucleotides useful as reagents in diagnostic assays for HCV.

For example, purified and isolated oligonucleotide sequences derived from the consensus sequences shown in Figures 6A-6K may be useful as genotype-specific primers and probes since these oligonucleotides can hybridize specifically to the nucleotide sequence of the core gene of HCV isolates belonging to a given genotype. Examples of regions of the consensus sequence of the core gene of a given genotype from which primers specific for that genotype may be deduced include but are not limited to, the nucleotide domains shown below for each

- 39 -

- ° genotype. The sequence in which the indicated nucleotide domains are found are indicated in parentheses to the right of each genotype.

Genotype 1 (Consensus Sequence of Figure 6C)

- 427-466, 444-483, 447-486 (5'-3', sense)  
5 505-466, 522-483, 525-486 (5'-3', antisense)

Genotype 1a (Consensus Sequence of Figure 6A)

- 141-180, 279-318 (5'-3', sense)  
219-180, 246-207 (5'-3', antisense)

10

Genotype 1b (Consensus Sequence of Figure 6B)

- 67-106, 127-186, 234-273 (5'-3', sense)  
144-106, 225-186, 311-272, 312-273 (5'-3', antisense)

15

Genotype 2 (Consensus Sequence of Figure 6F)

- 153-192, 162-201, 164-203, 168-207, 171-210, 182-221, 192-231, 193-232, 302-341 (5'-3', sense)  
231-192, 240-201, 242-203, 246-207, 249-210, 260-221, 270-231, 271-232, 380-341 (5'-3', antisense)

20

Genotype III/2a (Consensus Sequence of Figure 6D)

- 276-315, 306-355 (5'-3', sense)  
309-270, 354-315, 394-355, 571-532 (5'-3', antisense)

25

Genotype IV/2b (Consensus Sequence of Figure 6E)

- 6-45, 135-174, 177-216, 309-348, 337-376, 375-414, 501-540 (5'-3', sense)  
84-45, 213-174, 255-216, 387-348, 415-376, 453-414, 571-532, 573-540 (5'-3', antisense)

30

Genotype 2c (SEQ ID NO:134)

- 194-233, 273-312, 279-318, 417-456, 423-462, 504-543, 505-544, 517-556 (5'-3', sense)  
272-233, 351-312, 354-315, 357-318, 450-411, 495-456, 501-462, 573-543, 556-573 (5'-3', antisense)  
35

- 40 -

° Genotype 3 or Genotype V/3a (Consensus Sequence of Figure 6G)

8-47, 45-84, 68-107, 87-126, 88-127, 90-129, 111-150, 142-181, 173-212, 177-216, 261-300,

276-315, 452-491, 520-559, 521-560, 529-568, 532-571, 533-

5 572. (5'-3', sense)

86-47, 123-84, 146-107, 165-126, 186-147, 189-150, 219-180, 250-211, 251-212, 255-216,

339-300, 530-491, 573-543, 573-557, 573-559, 573-560. (5'-3', antisense)

10

Genotype 4 (Consensus Sequence of Figure 6H)

20-59 (5'-3', sense)

97-58, 98-59 (5'-3', antisense)

15 Genotype 4a (SEQ ID NO:139)

111-150, 150-189, 174-213, 183-222, 192-231, 261-300, 376-415, 396-435, 531-570 (5'-3', sense)

186-147, 252-213, 270 -231, 339-300, 454-415 (5'-3', antisense)

20

Genotype 4b (SEQ ID NO:141)

27-66, 30-69, 106-145, 271-310, 433-472, 447-486, 453-492 (5'-3', sense)

105-66, 183-144, 184-145, 345-306, 348-309, 349-310, 468-

25 429, 510-471, 522-483, 570-531 (5'-3', antisense)

Genotype 4c (SEQ ID NO:143)

174-213, 180-219, 207-246, 231-270 (5'-3', sense)

249-210, 252-213, 258-219, 309-270, 504-465 (5'-3',

30

antisense)

Genotype 4d (SEQ ID NO:145)

173-212, 188-327, 430-469 (5'-3', sense)

248-209, 249-210, 250-211, 251-212, 366-327, 508-469 (5'-3', antisense)

35

- 41 -

° Genotype 4e (SEQ ID NO:142)  
 160-199, 267-306, 287-326, 288-327, 524-564 (5'-3', sense)  
 238-199, 345-306, 365-326, 216-177, 522-483 (5'-3',  
 antisense)

5 Genotype 4f (SEQ ID NO:140)  
 18-57, 36-75, 228-267, 396-435 (5'-3', sense)  
 96-57, 114-75, 306-267 (5'-3', antisense)

Genotype 5 or 5a (Consensus Sequence of Figure 6I)  
 10 176-215, 177-216, 181-220, 195-234, 221-260, 252-291, 255-  
 294, 396-435, 435-474, 447-486, 498-537 (5'-3', sense)  
 254-215, 299-260, 310-271, 330-291, 333-294, 354-315, 464-  
 425, 471-432, 483-444, 570-531 (5'-3', antisense)

15 Genotype 6 or 6a (SEQ ID NO:154)  
 20-59, 136-175, 156-195, 159-198, 175-214, 185-224, 277-  
 316, 278-317, 312-351, 348-387, 405-444, 406-445, 407-446,  
 408-447, 411-450, 432-471, 433-472, 435-474, 522-561 (5'-  
 3', sense).  
 20 98-59, 214-175, 234-195, 237-198, 253-214, 262-223, 263-  
 224, 354-315, 355-316, 382-343, 390-351, 426-387, 468-429,  
 483-444, 484-445, 485-446, 486-447, 489-450, 510-471, 511-  
 472, 513-474 (5'-3', antisense)

25 Such nucleotide domains may range from about 15  
 to about 100 bases in length with a more preferred range  
 being about 30 to about 60 bases in length.

In an alternative embodiment, universal primers  
 able to hybridize to the nucleotide sequences of the core  
 gene of HCV isolates belonging to all of the genotypes  
 30 disclosed herein may be deduced from universally conserved  
 nucleotide domains of the consensus sequence shown in  
 Figures 6J and 6K. Examples of such nucleotide domains  
 include, but are not limited to, those shown below:

35 nucleotides 1-20, 1-25, 1-26, 1-27, 1-33, 50-89,  
 51-90, 52-91, 53-92, 61-100, 62-101, 77-116, 78-117, 79-

- 42 -

° 118, 80-119, 81-120, 82-121, 83-122, 84-123, 85-124, 86-  
 125, 97-136, 98-137, 99-138, 100-139, 101-140, 102-141,  
 329-368, 330-369, 331-370, 332-371, 354-393, 355-394, 356-  
 395, 362-401, 363-402, 364-403, 365-404, 369-408, 442-481,  
 443-482, 457-496, 458-497, 475-514, 476-515, 477-516 (5'-  
 5 3, sense); and  
           nucleotides 40-1, 41-2, 42-3, 43-4, 51-12, 52-13,  
 55-16, 56-17, 57-18, 58-19, 61-22, 62-23, 63-24,  
 64-25, 70-31, 124-85, 125-86, 126-87, 127-88, 128-89, 129-  
 90, 136-97, 137-98, 138-99,  
 10 149-110, 150-111, 151-112, 152-113, 153-114, 154-115, 155-  
 116, 156-117, 157-118, 158-119, 159-120, 170-131, 171-132,  
 172-133, 173-134, 174-135, 175-136, 403-364, 405-365, 406-  
 366, 406-367, 430-391, 431-392, 432-393, 436-397, 437-398,  
 438-399, 439-400, 517-478, 518-479, 519-480, 532-493, 533-  
 15 494, 550-511, 551-512 (5'-3', antisense)

Those skilled in the art would readily understand  
 that the term "antisense" as used herein refers to primer  
 sequences which are the complementary sequence of the  
 indicated consensus sequence or SEQ ID NO:. Further,  
 20 provided with the above examples of regions of the  
 consensus sequences or indicated SEQ ID NOS: from which to  
 deduce universal and genotype-specific primers, those  
 skilled in the art would readily be able to select pairs of  
 primers, one sense and one antisense, which would be useful  
 25 in the detection of HCV genotypes via the PCR methods  
 described herein.

In yet another embodiment, the sequences shown in  
 SEQ ID NO.:103-154 and the resultant consensus sequences  
 produced by alignment of these SEQ ID NOS as shown in  
 30 Figures 6A-6K may also be useful in the design of  
 hybridization probes specific for a given HCV genotype.  
 Examples of nucleotide domains of the consensus sequence or  
 SEQ ID NO of a given genotype from which genotype-specific  
 hybridization probes may be deduced include, but are not  
 35 limited to, those shown below where the sequence from which



- 43 -

- ° the domains are found is indicated in parentheses to the right of each genotype.

	<u>Genotype</u>	<u>Position</u>
	1a (Consensus sequence of Figure 6A)	50-85
		155-205
		207-277
5		281-333
		429-477
		530-573
	1b (Consensus sequence of Figure 6B)	81-131
		159-225
		252-318
10		411-472
		530-573
	2a (Consensus sequence of Figure 6D)	35-75
		200-276
		290-340
		330-380
		410-472
15		530-573
	2b (Consensus sequence of Figure 6E)	20-70
		149-199
		191-241
		240-285
		261-318
		323-373
20		351-401
		389-439
		429-477
		530-573
	2c (SEQ ID NO:134)	208-258
		230-276
		290-345
25		411-460
		430-490
		530-573
	3a (Consensus sequence of Figure 6G)	1-50
		40-100
		100-160
30		145-190
		190-240
		275-325
		411-455
		466-516
		530-573
35	4a (SEQ ID NO:139)	35-85
		145-195

- 44 -

°		200-250 255-305 341-390 390-440 530-573
5	4b (SEQ ID NO:141)	35-85 120-170 180-225 230-275 285-335 405-455 462-492 530-573
10	4c (SEQ ID NO:143)	35-85 190-246 245-295 282-318 372-415 440-480 530-573
15	4d (SEQ ID NO:145)	35-85 187-237 302-352 405-455 444-494 530-573
20	4e (SEQ ID NO:142)	35-85 57-84 174-224 230-275 290-340 422-472 530-573
25	4f (SEQ ID NO:140)	35-85 174-224 242-292 290-340 422-472 530-573
30	5a (Consensus sequence of Figure 6I)	180-234 265-315 315-355 420-486 530-573
35	6a (SEQ ID NO:154)	34-84 150-200 180-230

- 45 -

°		230-290
		291-333
		341-395
		429-490
		530-573
	1 (Consensus sequence of Figure 6C)	192-241
5		435-495
	2 (Consensus sequence of Figure 6F)	186-240
		320-360
		440-475
	4 (Consensus sequence of Figure 6H)	40-80

10           In yet another embodiment, universal hybridization probes may be derived from the consensus sequences shown in Figures 6J and 6K. Examples of nucleotide domains of the consensus sequences shown in Figure 6J and 6K from which universal hybridization probes  
15 may be derived include, but are not limited to, 1-33; 85-141; 364-408; 478-516.

The oligonucleotides of this invention can be synthesized using any of the known methods of oligonucleotide synthesis (e.g., the phosphodiester method  
20 of Agarwal et al. 1972, Agnew. Chem. Int. Ed. Engl. 11:451, the phosphotriester method of Hsiung et al. 1979, Nucleic Acids Res 6:1371, or the automated diethylphosphoramidite method of Baeucage et al. 1981, Tetrahedron Letters 22:1859-1862), or they can be isolated fragments of  
25 naturally occurring or cloned DNA. In addition, those skilled in the art would be aware that oligonucleotides can be synthesized by automated instruments sold by a variety of manufacturers or can be commercially custom ordered and prepared. In a preferred embodiment, the oligonucleotides  
30 of the present invention are synthetic oligonucleotides. The oligonucleotides of the present invention may range from about 15 to about 100 nucleotides; with the preferred sizes being about 20 to about 60 nucleotides; a more preferred size being about 25 to about 50 nucleotides; and  
35 a most preferred size being about 30 to about 40

- 46 -

- ° nucleotides.

The present invention also relates to methods for detecting the presence of HCV in a mammal, said methods comprising analyzing the RNA of a mammal for the presence of hepatitis C virus.

- 5           The RNA to be analyzed can be isolated from serum, liver, saliva, lymphocytes or other mononuclear cells as viral RNA, whole cell RNA or as poly(A)<sup>+</sup> RNA. Whole cell RNA can be isolated by methods known to those skilled in the art. Such methods include extraction of RNA  
10 by differential precipitation (Birnbiom, H.C. (1988) Nucleic Acids Res., 16:1487-1497), extraction of RNA by organic solvents (Chomczynski, P. et al. (1987) Anal. Biochem., 162:156-159) and extraction of RNA with strong denaturants (Chirgwin, J.M. et al. (1979) Biochemistry,  
15 18:5294-5299). Poly(A)<sup>+</sup> RNA can be selected from whole cell RNA by affinity chromatography on oligo-d(T) columns (Aviv, H. et al. (1972) Proc. Natl. Acad. Sci., 69:1408-1412). A preferred method of isolating RNA is extraction of viral RNA by the guanidinium-phenol-chloroform method of Bukh et  
20 al. (1992a).

- The methods for analyzing the RNA for the presence of HCV include Northern blotting (Alwine, J.C. et al. (1977) Proc. Natl. Acad. Sci., 74:5350-5354), dot and slot hybridization (Kafatos, F.C. et al. (1979) Nucleic  
25 Acids Res., 7:1541-1522), filter hybridization (Hollander, M.C. et al. (1990) Biotechniques; 9:174-179), RNase protection (Sambrook, J. et al. (1989) in "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Press, Plainview, NY) and reverse-transcription polymerase chain  
30 reaction (RT-PCR) (Watson, J.D. et al. (1992) in "Recombinant DNA" Second Edition, W.H. Freeman and Company, New York).

- A preferred method for analyzing the RNA is RT-PCR. In this method, the RNA can be reverse transcribed to  
35 first strand cDNA using a primer or primers derived from

- 47 -

° the nucleotide sequences shown in SEQ ID NOs:1-51 or SEQ ID NOs:103-154 or sequences complementary to those described. Once the cDNAs are synthesized, PCR amplification is carried out using pairs of primers designed to hybridize with sequences in the HCV E1 or core cDNA which are an appropriate distance apart (at least about 50 nucleotides) to permit amplification of the cDNA and subsequent detection of the amplification product. Alternatively, one can amplify both E1 and core cDNA sequences by using a primer pair where one primer hybridizes with the E1 cDNA sequence and the other primer hybridizes with the core cDNA sequence. Each primer of a pair is a single-stranded oligonucleotide of about 20 to about 60 bases in length with a more preferred range being about 30 to about 50 bases in length where one primer (the "upstream" primer) is complementary to the original RNA and the second primer (the "downstream" primer) is complementary to the first strand of cDNA generated by reverse transcription of the RNA. The target sequence is generally about 100 to about 300 base pairs long but can be as large as 500-1500 base pairs. Optimization of the amplification reaction to obtain sufficiently specific hybridization to the nucleotide sequence of interest (either E1 or core or both E1 and core) is well within the skill in the art and is preferably achieved by adjusting the annealing temperature.

25 In one embodiment, the primer pairs selected to amplify E1 and core cDNAs are universal primers. By "universal", as used to describe primers throughout the claims and specification, is meant those primer pairs which can amplify E1 and/or core gene fragments derived from an HCV isolate belonging to any one of the genotypes of HCV described herein. Purified and isolated universal primers for E1 cDNAs are used in Example 1 of the present invention and are shown as SEQ ID NOs:207-212 where SEQ ID NOs:207 and 208 represent one pair of primers, SEQ ID NOs:209 and 210 represent a second pair of primers and SEQ ID NOs:211-

- 48 -

212 represent a third pair of primers. Nucleotide domains of the consensus sequence shown in Figure 6J from which universal primers for core cDNAs may be deduced have previously been disclosed within the present specification. Alternatively, a universal primer for E1 cDNA sequence and a universal primer for core cDNA sequence may be used as a universal primer pair to amplify both E1 and core cDNAs.

In an alternative embodiment, primer pairs selected to amplify E1 and/or core cDNAs are genotype-specific primers. In the present invention, genotype-specific primer pairs can readily be derived from the following genotype-specific E1 nucleotide domains: nucleotides 197-238 and 450-480 of the consensus sequence of genotype I/1a shown in Figure 1A; nucleotides 197-238 and 450-480 of the consensus sequence of genotype II/1b shown in Figure 1B; nucleotides 199-238 and 438-480 of the consensus sequence of genotype III/2a shown in Figure C; nucleotides 124-177 and 450-480 of the consensus sequence of genotype IV/2b shown in Figure 1D; nucleotides 124-177, 193-238 and 436-480 of SEQ ID NO:34 (genotype 2C); nucleotides 168-207, 294-339 and 406-480 of the consensus sequence of genotype V/3a shown in Figure 1E; nucleotides 145-183 and 439-480 of SEQ ID NO:40 (genotype 4a); nucleotides 168-207 and 432-480 of SEQ ID NO:41 (genotype 4b); nucleotides 130-183 and 450-480 of the consensus sequence of genotype 4c shown in Figure 1F; nucleotides 130-183 and 450-480 of SEQ ID NO:44 (genotype 4d); nucleotides 166-208 and 437-480 of the consensus sequence of genotype 5a shown in Figure 1b and nucleotides 168-207, 216-252 and 429-480 of SEQ ID NO:51 (genotype 6a). Genotype-specific HCV core nucleotide domains from which genotype-specific primers may be deduced have previously been described herein. Those skilled in the art would readily appreciate that in a pair of genotype-specific primers, each primer is derived from different nucleotide domains specific for a given genotype. Also, it is

- 49 -

° understood by those skilled in the art that each pair of primers comprises one primer which is complementary to the original viral RNA and the other which is complementary to the first strand of cDNA generated by reverse transcription of the viral RNA. For example, in a pair of genotype-specific primers for genotype 4b, one primer would have a nucleotide sequence derived from region 168-207 of SEQ ID NO:40 and the other primer would have a nucleotide sequence which is the complement of region 432-480 of SEQ ID NO:40. One skilled in the art would readily recognize that such genotype-specific domains would also be useful in designing oligonucleotides for use as genotype-specific hybridization probes. Indeed, genotype-specific hybridization probes deduced from the E1 and core sequences of the present invention have been previously disclosed herein.

15           The amplification products of PCR can be detected either directly or indirectly. In one embodiment, direct detection of the amplification products is carried out via labelling of primer pairs. Labels suitable for labelling the primers of the present invention are known to one skilled in the art and include radioactive labels, biotin, 20           avidin, enzymes and fluorescent molecules. The derived labels can be incorporated into the primers prior to performing the amplification reaction. A preferred labelling procedure utilizes radiolabeled ATP and T4 25           polynucleotide kinase (Sambrook, J. et al. (1989) in "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Press, Plainview, NY). Alternatively, the desired label can be incorporated into the primer extension products during the amplification reaction in the form of one or more labelled dNTPs. In the present invention, the 30           labelled amplified PCR products can be detected by agarose gel electrophoresis followed by ethidium bromide staining and visualization under ultraviolet light or via direct sequencing of the PCR-products. Thus, in one embodiment, 35           the present invention relates to a method for determining

- 50 -

- ° the genotyp of a hepatitis C virus present in a mammal where said method comprises: amplifying RNA of a mammal via RT-PCR using labelled genotype-specific primers for the amplification step of the cDNA produced by reverse transcription.

5           In yet another embodiment, unlabelled amplification products can be detected via hybridization with labelled nucleic acid probes radioactively labelled or, labelled with biotin, in methods known to one skilled in the art such as dot and slot blot hybridization  
10       (Kafatos, F.C. et al. (1979) or filter hybridization (Hollander, M.C. et al. (1990)).

          In one embodiment, the nucleic acid sequences used as probes are selected from, and substantially homologous to, SEQ ID NOS:1-51 and/or SEQ ID NOS:103-154.  
15       Such probes are useful as universal probes in that they can detect PCR-amplification products of E1 and/or core cDNAs of an HCV isolate belonging to any of the HCV genotypes disclosed herein. The size of these probes can range from about 200 to about 500 nucleotides. In an alternative  
20       embodiment, the sequence alignments shown in Figures 1A-1H and 6A-6J may be used to design oligonucleotides useful as universal hybridization probes. Examples of core and envelope nucleotide domains from which such universal oligonucleotides may be deduced are disclosed herein.

25           In yet another embodiment, the present invention relates to a method for determining the genotype of a hepatitis C virus present in a mammal where said method comprises:

- 30           (a) amplifying RNA of a mammal via RT-PCR to produce amplification products;
- (b) contacting said products with at least one genotype-specific oligonucleotide; and
- (c) detecting complexes of said products which bind to said oligonucleotide(s).

35           In this method, one embodiment of said



- 51 -

° amplification step is carried out using the universal primers for E1 or core cDNAs as disclosed above. In step (b) of this method, the genotype-specific sequences used as probes may be deduced from the genotype-specific E1 and core nucleotide domains disclosed herein. These probes are  
5 useful in specifically detecting PCR-amplification products of E1 or core cDNAs of HCV isolates belonging to one of the HCV genotypes disclosed herein. In a preferred embodiment, these probes are used alone or in combination with other probes specific to the same genotype.

10 For example, a probe having a sequence according to SEQ ID NO:213 can be used alone or in combination with a probe having a sequence according to SEQ ID NO:214. The probes used in this method can range in size from about 15 to about 100 nucleotides with a more preferred range being  
15 about 30 to about 70 nucleotides. Such probes can be synthesized as described earlier.

In an alternative embodiment, the genotype of the amplification product of step (a) may be determined by using the nucleic acid sequences shown in SEQ ID NOs: 1-51  
20 and 103-154 as probes (Delwart, E. et al. (1993)) Science, 262: 1257-1261). Probes utilized in the method of Delwart et al. may range in size from about 100 to about 1,000 nucleotides with a more preferred probe size being about  
25 200 to about 800 base pairs and a most preferred probe size being about 300 to about 700 nucleotides.

The nucleic acid sequence used as a probe to detect PCR amplification products of the present invention can be labeled in single-stranded or double-stranded form. Labelling of the nucleic acid sequence can be carried out  
30 by techniques known to one skilled in the art. Such labelling techniques can include radiolabels and enzymes (Sambrook, J. et al. (1989) in "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In addition, there are known non-radioactive  
35 techniques for signal amplification including methods for

- 52 -

- ° attaching chemical moieties to pyrimidine and purine rings  
(Dale, R.N.K. et al. (1973) Proc. Natl. Acad. Sci.,  
70:2238-2242; Heck, R.F. (1968) S. Am. Chem. Soc., 90:5518-  
5523), methods which allow detection by chemiluminescence  
(Barton, S.K. et al. (1992) J. Am. Chem. Soc., 114:8736-  
5 8740) and methods utilizing biotinylated nucleic acid  
probes (Johnson, T.K. et al. (1983) Anal. Biochem.,  
133:126-131; Erickson, P.F. et al. (1982) J. of Immunology  
Methods, 51:241-249; Matthaei, F.S. et al. (1986) Anal.  
Biochem., 157:123-128) and methods which allow detection by  
10 fluorescence using commercially available products.

The present invention also relates to computer  
analysis of the amino acid sequences shown in SEQ ID  
NOs:52-102 by the program GENALIGN. This analysis groups  
the 51 amino acid sequences shown in SEQ ID NOs:52-102 into  
15 twelve genotypes based upon the degree of variation of the  
amino acid sequences. For the purposes of the present  
invention, the amino acid sequence identity of E1 amino  
acid sequences of the same genotype ranges from about 85%  
to about 100% whereas the identity of E1 amino acid  
20 sequences of different genotypes ranges from about 45% to  
about 80%.

The grouping of SEQ ID NOs:52-102 into twelve HCV  
genotypes is shown below:

	<u>SEQ ID NOs:</u>	<u>Genotypes</u>
25	52-59	I/1a
	60-76	II/1b
	77-80	III/2a
	81-84	IV/2b
	85	2c
	86-90	V/3a
30	91	4a
	92	4b
	93-94	4c
	95	4d
	96-101	5a
	102	6a

35 For those genotypes containing more than one E1

- 53 -

° amino acid sequence, computer alignment of the constituent sequences of each genotype was conducted using the computer program GENALIGN in order to produce a consensus sequence for each genotype. These alignments and their resultant consensus sequences are shown in Figures 2A-G for the seven  
 5 genotypes (I/1a, II/1b, III/2a, IV/2b, V/3a, 4c and 5a) which comprise more than one sequence. Further alignment of the consensus sequences shown in Figures 2A-G with the amino acid sequences of SEQ ID NO:85 (genotype 2c); SEQ ID NO:91 (genotype 4a); SEQ ID NO:92 (genotype 4b); SEQ ID  
 10 NO:95 (genotype 4d) and SEQ ID NO:102 (genotype 6a) to produce a consensus amino acid sequence for all twelve genotypes is shown in Figure 2H. The multiple alignment of E1 amino acid sequences shown in Figures 2A-H produces consensus sequences which serve to highlight regions of  
 15 homology and non-homology between E1 amino acid sequences of the same genotype and of different genotypes and hence, these alignments can readily be used by those skilled in the art to design peptides useful in assays and vaccines for the diagnosis and prevention of HCV infection.

20 In another embodiment, the computer analysis of SEQ ID NOS: 155-206 by the probe genome results in distribution of the 52 HCV core sequences into 14 genotypes based upon identification of genotype-specific amino acid sequences.

25 The grouping of SEQ ID NOS: 155-206 into 14 HCV genotypes is shown below:

<u>SEQ ID NOS:</u>	<u>Genotypes</u>
155-160	I/1a
161-176	II/1b
30 177-180	III/2a
181-185	IV/2b
186	2c
187-190	V/3a
191	4a
193	4b
195	4c
35 196	4c
197	4d

- 54 -

°	194	4e
	192	4f
	198-205	5a
	206	6a

These fourteen genotypes can be further grouped into six major genotypes designated genotypes 1-6 as described earlier for the core nucleotide sequences of the present application. Computer alignment of the amino acid sequences disclosed in SEQ ID NOS: 155-206 are shown in figures 7A-7J. As with the multiple alignments of the E-1 amino acid sequences, the consensus sequences shown in figure 7A-7J serve to highlight regions of homology and nonhomology between core amino acid sequences of the same genotype and of different genotypes and hence, these alignments can readily be used by those skilled in the art to design peptides useful in assays and vaccines for the diagnosis and prevention of HCV infection.

Examples of purified and isolated peptides deduced from the alignments shown in Figures 2A-2H include, but are not limited to, SEQ ID NOS:240-263 wherein these peptides are derived from two regions of the amino acid sequences shown in Figures 2A-H, amino acids 48-80 and amino acids 138-160. The peptides shown in SEQ ID NOS. 240-263 are useful as genotype-specific diagnostic reagents since they are capable of detecting an immune response specific to HCV isolates belonging to a single genotype. The genotype-specificity of the peptides shown in SEQ ID NOS:240-263 are as follows: SEQ ID NOS:240 and 252 are specific for genotype IV/2b; SEQ ID NOS:241 and 253 are specific for genotype 2c; SEQ ID NOS:242 and 254 are specific for genotype III/2a; SEQ ID NOS:243 and 255 are specific for genotype V/a; SEQ ID NOS:244 and 256 are specific for genotype II/1b; SEQ ID NOS:245 and 257 are specific for genotype I/1a; SEQ ID NOS:246 and 258 are specific for genotype 4a; SEQ ID NOS:247 and 259 are specific for genotype 4c; SEQ ID NOS:248 and 260 are specific for genotype 4d; SEQ ID NOS:249 and 261 are

- 55 -

° specific for genotype 4b; SEQ ID NOs:250 and 262 are specific for genotype 5a and SEQ ID NOs:251 and 263 are specific for genotype 6a. In SEQ ID NO:240, Xaa at position 22 is a residue of Ala or Thr, Xaa at position 24 is a residue of Val or Ile, Xaa at position 26 is a residue of Val or Met; in SEQ ID NO:242, Xaa at position 5 is a Ser or Thr residue, Xaa at position 11 is an Arg or Gln residue, Xaa at position 12 is an Arg or Gln residue; in SEQ ID NO:243, Xaa at position 3 is a Pro or Ser residue, Xaa at position 33 is a Leu or Met residue; in SEQ ID NO:244, Xaa at position 5 is a Thr or Ala residue, Xaa at position 13 is a Gly, Ala, Ser, Val or Thr residue, Xaa at position 14 is a Ser, Thr or Asn residue, Xaa at position 15 is a Val or Ile residue, Xaa at position 16 is a Pro or Ser residue, Xaa at position 18 is a Thr or Lys residue, Xaa at position 19 is a Thr or Ala residue, Xaa at position 22 is an Arg or His residue, Xaa at position 32 is an Ala, Val or Thr residue; in SEQ ID NO:245, Xaa at position 3 is an Ala or Pro residue, Xaa at position 4 is a Val or Met residue, Xaa at position 5 is a Thr or Ala residue, Xaa at position 17 is a Thr or Ala residue, Xaa at position 18 is a Thr or Ala residue, Xaa at position 23 is a His or Tyr residue; in SEQ ID NO:247, Xaa at position 10 is a Val or Ala residue, Xaa at position 11 is a Ser or Pro residue, Xaa at position 18 is an Asp or Glu residue Xaa at position 20 is a Leu or Ile residue; in SEQ ID NO:250, Xaa at position 3 is a Gln or His residue, Xaa at position 12 is an Asn, Ser or Thr residue, Xaa at position 13 is a Leu or Phe residue, Xaa at position 23 is an Ala or Val residue; in SEQ ID NO:252, Xaa at position 16 is a Val or Ala residue, Xaa at position 18 is a Glu or Gln residue; in SEQ ID NO:254, Xaa at position 2 is an Ala or Thr residue, Xaa at position 4 is a Met or Leu residue, Xaa at position 9 is an Ala or Val residue, Xaa at position 17 is an Ile or Leu residue, Xaa at position 20 is an Ile or Val residue, Xaa at position 21 is a Ser or Gly residue; in SEQ ID NO:151,

- 56 -

- ° Xaa at position 9 is a Val or Ile residue, Xaa at position 16 is a Leu or Val residue, Xaa at position 20 is an Ile or Leu residue; in SEQ ID NO:256, Xaa at position 2 is an Ala or Thr residue, Xaa at position 6 is a Val or Leu residue, Xaa at position 12 is an Ile or Leu residue, Xaa at position 16 is a Val or Ile residue, Xaa at position 17 is a Val, Leu or Met residue, Xaa at position 19 is a Met or Val residue, Xaa at position 21 is an Ala or Thr residue; in SEQ ID NO:257, Xaa at position 2 is a Thr or Ala residue, Xaa at position 6 is a Val, Ile or Met residue, Xaa at position 12 is an Ile or Val residue, Xaa at position 16 is a Ile or Val residue; in SEQ ID NO:155, Xaa at position 5 is a Leu or Val residue, Xaa at position 21 is a Thr or Ala residue; in SEQ ID NO:262, Xaa at position 1 is a Thr or Ala residue, Xaa at position 5 is a Val or Leu residue, Xaa at position 9 is a Leu, Met or Val residue, Xaa at position 23 is a Gly or Ala residue.

Examples of core amino acid domains from which genotype-specific peptides may be deduced, include but are not limited to, those shown below where the sequence in which the indicated domains are found is given in parentheses to the right of each genotype:

	<u>Genotype</u>	<u>Amino Acid Domains</u>
	1a (consensus sequence of Figure 7A)	67-78
	1b (consensus sequence of Figure 7B)	67-78
	2 (consensus sequence of Figure 7F)	66-81
25		110-119
	2a (consensus sequence of Figure 7D)	67-78
		115-125
	2b (consensus sequence of Figure 7E)	67-78
		123-133
	2c (SEQ ID NO:186)	67-78
		75-81
		184-191
30	3a (consensus sequence of Figure 7G)	8-22
		32-46
		67-78
		158-170
		180-191
	4 (consensus sequence of Figure 7H)	14-23
	4a (SEQ ID NO:191)	67-78
35	4b (SEQ ID NO:193)	45-57

- 57 -

°		67-78
	4c (SEQ ID NO:195)	67-78
	4d (SEQ ID NO:197)	67-78
	4e (SEQ ID NO:194)	67-78
	4f (SEQ ID NO:192)	67-78
	5a (consensus sequence of Figure 7J)	67-78
	6a (SEQ ID NO:206)	67-78
5		101-108
		144-155
		157-163

Those skilled in the art would be aware that the peptides of the present invention or analogs thereof can be synthesized by automated instruments sold by a variety of manufacturers or can be commercially custom-ordered and prepared. The term analog has been described earlier in the specification and for purposes of describing the peptides of the present invention, analogs can further include branched, cyclic or other non-linear arrangements of the peptide sequences of the present invention.

Alternatively, peptides can be expressed from nucleic acid sequences where such sequences can be DNA, cDNA, RNA or any variant thereof which is capable of directing protein synthesis. In one embodiment, restriction digest fragments containing a coding sequence for a peptide can be inserted into a suitable expression vector that functions in prokaryotic or eukaryotic cells. Such restriction digest fragments may be obtained from clones isolated from prokaryotic or eukaryotic sources which encode the peptide sequence.

Suitable expression vectors and methods of isolating clones encoding the peptide sequences of the present invention have previously been described. In yet another embodiment, an oligonucleotide capable of directing host organism synthesis of the given peptide may be synthesized and inserted into the expression vector.

The preferred size of the peptides of the present invention is from about 8 to about 100 amino acids in length when the peptides are chemically synthesized with a more preferred size being about 8 to about 30 amino acids

- 58 -

° and a most preferred size being about 10 to about 20 amino acids in length. For recombinantly expressed peptides, the size may range from about 20 to about 190 amino acids in length with a more preferred size being about 70 amino acids.

5           The present invention further relates to the use of genotype-specific peptides in methods of detecting antibodies against a specific genotype of HCV in biological samples. In one embodiment, at least one genotype-specific peptide deduced from a genotype-specific core or E1 amino  
10 acid domain may be used in any of immunoassays described herein to detect antibodies specific for a single genotyp of HCV. In another embodiment, at least one genotype-specific peptide deduced from a genotype-specific core nucleotide domain and at least one genotype-specific  
15 peptide deduced from an E1 amino acid domain may be used in an immunoassay to detect antibodies against a single genotype of HCV. A preferred immunoassay is ELISA.

          It is understood by those skilled in the art that the diagnostic assays described herein using genotype-specific oligonucleotides or genotype-specific peptides can  
20 be useful in assisting one skilled in the art to choose a course of therapy for the HCV-infected individual.

          In an alternative embodiment, a mixture of genotype-specific peptides can be used in an immunoassay to  
25 detect antibodies against multiple genotypes of HCV disclosed herein. For example, a mixture of genotype-specific peptides deduced from E1 amino acid sequences may comprise at least one peptide selected from SEQ ID NOs:244-245 and 256-257; one peptide selected from SEQ ID NOs:240,  
30 242, 252 and 254; one peptide selected from SEQ ID NOs:246-249 and 258-261; one peptide selected from SEQ ID NOs:250 and 262; one peptide selected from SEQ ID NOs:243 and 255; one peptide selected from SEQ ID NOs:242 and 254 and one peptide selected from SEQ ID NOs:244 and 263. In a  
35 preferred embodiment, the peptides of the present invention



- 59 -

- ° can be used in an ELISA assay as described previously for recombinant E1 and core proteins.

In an alternative embodiment, the peptide(s) utilized in an immunoassay to detect all the genotypes of HCV disclosed herein may be a universal peptide deduced from universally conserved amino acid domains of the E1 or core proteins disclosed herein.

Examples of universally conserved core amino acid domains within the consensus sequence shown in Figure 7J from which universal peptides may be deduced include, but are not limited to amino acid domains 23-35, 53-66, 93-108, 122-138, 150-156, and 165-181 of the consensus sequence. Examples of universally conserved E1 amino acid domains within the HCV E1 protein are located within the consensus sequence for the 51 HCV E1 proteins shown in Figure 2H of the present application. Examples of universally conserved domains within the consensus sequence shown in Figure 2H include, but are not limited to, amino acid domains 10-20, 111-120, and 124-137 of the consensus sequence. The universal peptides of the present invention may be used in an immunoassay to detect antibodies in patient sera specific for any of the genotypes of HCV disclosed herein.

The peptides of the present invention or analogs thereof may be prepared in the form of a kit, alone or in combinations with other reagents such as secondary antibodies, for use in immunoassay.

In another embodiment, the genotype-specific and universal peptides of the present invention may be used to produce antibodies that will react against HCV E1 or core proteins in immunoassays. In one embodiment, a genotype-specific E1 or core peptide can be used alone or in combination with other E1 or core peptides specific to the same genotype as immunogens to produce antibodies specific to HCV proteins of a single genotype.

In another embodiment, a mixture of peptides specific for different genotypes may be used to produce

- 60 -

° antibodies that will react with HCV proteins of any genotype disclosed herein. More preferably, antibodies reactive with HCV proteins of any genotype may be produced by immunizing an animal with universal peptide(s) of the present invention. Examples of immunoassays in which such antibodies could be utilized to detect HCV E1 and core proteins in biological samples include, but are not limited to, radioimmunoassays and ELISAs. Examples of biological samples in which HCV E1 and core proteins could be detected includes, but it is not limited to, serum, saliva and liver.

Of course, those skilled in the art would readily understand that the genotype-specific and universal peptides of the present invention and expression vectors containing nucleic acid sequence capable of directing host organism synthesis of these peptides could also be used as vaccines against hepatitis C. Formulations suitable for administering the peptide(s) and expression vectors of the present invention as immunogen, routes of administration, pharmaceutical compositions comprising the peptides expression vectors and so forth are the same as those previously described for recombinant E1 and core proteins.

The genotype-specific and universal peptides of the present invention and expression vectors containing nucleic acid sequence capable of direct host organism synthesis of these peptides may also be supplied in the form of a kit, alone, or in the form of a pharmaceutical composition as described above for recombinant E1 and core proteins.

Any articles or patents referenced herein are incorporated by reference. The following examples illustrate various aspects of the invention but are in no way intended to limit the scope thereof.

- 61 -

MATERIALS

Serum used in these examples was obtained from 84 anti-HCV positive individuals who were previously found to be positive for HCV RNA in a cDNA PCR assay with primer set a from the 5' NC region of the HCV genome (Bukh, J. et al. (1992 (b)) Proc. Natl. Acad. Sci. USA 89:4942-4946). These samples were from 12 countries: Denmark (DK); Dominican Republic (DR); Germany (D); Hong Kong (HK); India (IND); Sardinia, Italy (S); Peru (P); South Africa (SA); Sweden (SW); Taiwan (T); United States (US); and Zaire (Z).

Example 1

Identification of the cDNA Sequence  
of the E1 Gene of 51 Isolates of HCV via  
RT-PCR Analysis of Viral RNA Using Universal Primers

Viral RNA was extracted from 100  $\mu$ l of serum by the guanidinium-phenol-chloroform method and the final RNA solution was divided into 10 equal aliquots and stored at -80°C as described (Bukh, et al. (1992 (a))). The sequences of the synthetic oligonucleotides used in the RT-PCR assay, deduced from the sequence of HCV strain H-77 (Ogata, N. et al. (1991) Proc. Natl. Acad. Sci. USA 88:3392-3396), are shown as SEQ ID NOS:207-212. One aliquot of the final RNA solution, equivalent to 10  $\mu$ l of serum, was used for cDNA synthesis that was performed in a 20  $\mu$ l reaction mixture using avian myeloblastosis virus reverse transcriptase (Promega, Madison, WI) and SEQ ID NO:208 as a primer. The resulting cDNA was amplified in a "nested" PCR assay by Taq DNA polymerase (Amplitaq, Perkin-Elmer/Cetus) as described previously (Bukh et al. (1992a)) with primer set e (SEQ ID NOS:207-210). Precautions were taken to avoid contamination with exogenous HCV nucleic acid (Bukh et al. (1992a)), and negative controls (normal, uninfected serum) were interspersed between every test sample in both the RNA extraction and cDNA PCR procedures. No false positive results were observed in the analysis. In most instances,

- 62 -

° amplified DNA (first or second PCR products) was reamplified with primers SEQ ID NO:211 and SEQ ID NO:212 prior to sequencing since these two primers contained EcoR1 sites which would facilitate future cloning of the E1 gene. Amplified DNA was purified by gel electrophoresis followed by glass-milk extraction (Geneclean, BIO 101, LaJolla, CA) and both strands were sequenced directly by the dideoxynucleotide chain termination method (Bachman, B. et al. (1990) Nucl. Acids Res. 18:1309)) with phage T7 DNA polymerase (Sequenase, United States Biochemicals, Cleveland, OH), [ $\alpha$   $^{35}$ S]dATP (Amersham, Arlington Heights, IL) or [ $\alpha$   $^{33}$ P] dATP (Amersham or DuPont, Wilmington, DE) and sequencing primers. RNA extracted from serum containing HCV strain H-77, previously sequenced by Ogata, N. et al. (1991), was amplified with primer set e (SEQ ID NOS:207-210) and sequenced in parallel as a control. The nucleotide sequences of the envelope 1 (E1) gene of all 51 HCV isolates are shown as SEQ ID NOS:1 - 51. In all 51 HCV isolates, the E1 gene was exactly 576 nucleotides in length and did not have any in-frame stop codons.

#### Example 2

#### Computer Analysis of the Nucleotide and Deduced Amino Acid Sequences of the E1 Gene of 51 HCV Isolates

25 Multiple computer-generated alignments of the nucleotide (SEQ ID NOS:1-51, Figures 1A-H) and deduced amino acid sequences (SEQ ID NOS:52-102, Figures 2A-H) of the cDNAs of the 51 HCV isolates constructed using the computer program GENALIGN (Miller, R.H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:2057-2061) resulted in the 51 HCV isolates being divided into twelve genotypes based upon the degree of variation of the E1 gene sequence as shown in table 1.

35

Biochemistry: Butth *et al.*

Table 1. Percent nucleotide (nt) and amino acid (aa) sequence identity of the E1 gene among the 12 HCV genotypes.

	I/1a	II/1b	III/2a	IV/2b	2c	(V)/3a	4a	4b	4c	4d	5a	6a	nt:
I/1a	89.9-97.6	72.0-76.2	59.2-63.7	56.1-58.3	60.8-62.8	63.0-66.3	63.9-67.2	64.9-66.8	62.7-64.4	67.7-69.4	62.3-67.2	62.2-63.9	I/1a
II/1b	88.9-97.9	88.9-97.9	58.3-62.2	53.8-57.5	60.1-61.5	63.9-67.2	60.9-63.7	63.4-65.8	61.6-65.1	63.0-65.5	62.2-66.5	61.6-63.0	II/1b
III/2a	91.1-98.4	88.0-91.3	88.0-91.3	69.1-71.0	72.7-73.6	58.0-60.8	61.5-62.7	58.9-60.4	59.7-63.4	58.7-61.3	56.6-60.8	55.0-56.8	III/2a
IV/2b	75.5-80.7	90.1-97.9	89.1-92.7	92.7-95.0	67.5-68.9	56.3-58.3	58.9-60.8	56.4-57.6	57.1-59.9	57.5-59.0	53.5-56.6	53.6-55.2	IV/2b
2c	58.3-64.6	52.6-56.8	89.1-92.7	93.8-96.4	---	57.5-58.2	59.2	58.5	58.0-58.3	58.9	56.9-57.1	57.6	2c
(V)/3a	54.2-56.8	51.0-54.2	69.3-72.9	93.8-96.4	---	93.8-99.1	64.4-65.3	62.7-64.1	60.9-62.5	62.3-63.9	61.8-64.4	58.0-58.9	(V)/3a
4a	56.3-60.4	52.6-55.7	74.5-77.1	67.7-69.8	---	---	---	74.8	75.5-78.0	74.8	62.8-64.6	62.0	4a
4b	64.1-68.8	66.7-70.8	54.7-58.9	54.2-56.8	52.1-53.6	94.3-98.4	---	---	74.0-74.8	72.0	63.9-64.6	62.7	4b
4c	69.3-73.4	64.6-67.2	62.0-63.0	58.9-60.4	58.3	66.1-68.8	---	---	90.1	77.6-78.6	62.7-64.8	63.0-64.4	4c
4d	66.7-69.3	66.1-70.3	53.6-56.3	52.1-53.1	53.6	62.0-64.6	76.0	---	---	---	64.4-66.1	64.1	4d
5a	66.1-72.9	64.6-69.3	55.2-61.5	54.2-58.3	54.7-58.3	63.0-65.6	77.1-81.3	79.2-80.2	89.6	---	90.1-95.7	60.6-63.2	5a
6a	73.4-75.5	66.7-70.3	56.3-58.9	55.2-55.7	54.2	63.5-64.6	78.1	77.6	82.8	---	---	---	6a
	66.1-73.4	64.1-70.3	52.6-57.3	50.5-53.1	54.2-56.3	60.4-64.1	67.2-68.2	65.1-67.2	67.7-71.4	69.3-71.4	92.7-97.4	---	
	64.6-65.6	62.5-65.6	49.0-51.0	49.0-50.5	50.5	57.8-58.9	66.1	62.5	66.1-67.2	66.7	62.0-63.5	---	

Nucleotide sequences analyzed in compiling the above table are shown in SEQ ID NOs:1-51 while the amino acid sequences analyzed are shown in SEQ ID NOs:52-102. The grouping of SEQ ID NOs: into genotypes is previously described in the specification.

- 64 -

° The nucleotide and amino acid sequence identity of HCV isolates of the same genotype was in the range of 88.0-99.1% and 89.1-98.4%, respectively, whereas that of HCV isolates of different genotypes was in the range of 53.5-78.6% and 49.0-82.8%, respectively. The latter differences are similar to those found when comparing the envelope gene sequences of the various serotypes of the related flaviviruses, as well as other RNA viruses. When microheterogeneity in a sequence was observed, defined as more than one prominent nucleotide at a specific position, the nucleotide that was identical to that of the HCV prototype (HCV1, Choo et al. (1989)) was reported if possible. Alternatively, the nucleotide that was identical to the most closely related isolate is shown.

Analysis of the consensus sequence of the E1 protein of the 51 HCV isolates from this study demonstrated that a total of 60 (30.3%) of the 192 amino acids of the E1 protein were invariant among these isolates (Fig. 3). Most impressive, all 8 cysteine residues as well as 6 of 8 proline residues were invariant. The most abundant amino acids (e.g. alanine, valine and leucine) showed a very low degree of conservation. The consensus sequence of the E1 protein contained 5 potential N-linked glycosylation sites. Three sites at positions 209, 305 and 325 were maintained in all 51 HCV isolates. A site at position 196 was maintained in all isolates except the sole isolate of genotype 2c. Also, a site at position 234 was maintained in all isolates except one isolate of genotype I/1a, all four isolates of genotype IV/2b and the sole isolate of genotype 6a. Conversely, only genotype IV/2b isolates had a potential glycosylation site at position 233. Further analysis revealed a highly conserved amino acid domain (aa 302-328) in the E1 protein with 20 (74.1%) of 27 amino acids invariant among all 51 HCV isolates. It is possible that the 5' and 3' ends of this domain are conserved due to important cysteine residues and N-linked glycosylation

- 65 -

° sites. The central sequence, 5'-GHRMAWDMM-3' (aa 315-323), may be conserved due to additional functional constraints on the protein structure. Finally, although the amino acid sequence surrounding the putative E1 protein cleavage site was variable, an amino acid doublet (GV) at position 380 was invariant among all HCV isolates.

5 A dendrogram of the genetic relatedness of the E1 protein of selected HCV isolates representing the 12 genotypes is shown in Fig. 4. This dendrogram was constructed using the program CLUSTAL (Higgins, D.G. et al. 10 (1988) Gene, 73:237-244) and had a limit of 25 sequences. The scale showing percent identity was added based upon manual calculation. From the 51 HCV isolates for which the complete sequence of the E1 gene region was obtained, 25 isolates representing the twelve genotypes were selected 15 for analysis. This dendrogram in combination with the analysis of the E1 gene sequence of 51 HCV isolates in Table 1 demonstrates extensive heterogeneity of this important gene.

The worldwide distribution of the 12 genotypes among 74 HCV isolates is depicted in Fig. 5. The complete E1 gene sequence was determined in 51 of these HCV isolates (SEQ ID NOS:1-51), including 8 isolates of genotype I/1a, 17 isolates of genotype II/1b and 26 isolates comprising genotypes III/2a, IV/2b, 2c, 3a, 4a-4d, 5a and 6a. In the 25 remaining 23 isolates, all of genotypes I/1a and II/1b, the genotype assignment was based on a partial E1 gene sequence since they did not represent additional genotypes in any of the 12 countries. The number of isolates of a particular genotype is given in each of the 12 countries studied. Of 30 the twelve genotypes, genotypes I/1a and II/1b were the most common - accounting for 48 (65%) of the 74 isolates. Analysis of the E1 gene sequences available in the GenBank data base at the time of this study revealed that all 44 such sequences were of genotypes I/1a, II/1b, III/2a and 35 IV/2b. Thus, based upon E1 gene analysis, 8 new genotypes

- 66 -

° of HCV have been identified.

Also of interest, different HCV genotypes were frequently found in the same country, with the highest number of genotypes (five) being detected in Denmark. Of the twelve genotypes, genotypes I/1a, II/1b, III/2a, IV/2b and V/3a were widely distributed with genotype II/1b being identified in 11 of 12 countries studied (Zaire was the only exception). In addition, while genotypes I/1a and II/1b were predominant in the Americas, Europe and Asia, several new genotypes were predominant in Africa.

It was also found that genotypes I/1a, II/1b, III/2a, IV/2b and V/3a of HCV were widely distributed around the world, whereas genotypes 2c, 4a, 4b, 4d, 5a and 6a were identified only in discreet geographical regions. For example, the majority of isolates in South Africa comprised a new genotype (5a) and all isolates in Zaire comprised 3 new closely related genotypes (4a, 4b, 4c). These genotypes were not identified outside Africa.

### Example 3

#### Identification of the cDNA Sequence Of The Core Gene Of 52 Isolates Of HCV

Viral RNA extraction, cDNA synthesis and "nested" PCR were carried out as in Example 1. For the cDNA PCR assay HCV-specific synthetic oligonucleotides deduced from previously determined sequences that flank the C gene were used. Amplified DNA was purified by gel electrophoresis followed by glass-milk extraction as described in Example 1 or by electroelution and both strands were sequenced directly. In 44 of the 52 HCV isolates studied the procedures for direct sequencing described in Example 1 were utilized. For a number of the HCV isolates confirmatory sequencing was performed with the Applied Biosystems 373A automated DNA sequencer and 8 HCV isolates of genotype I/1a or II/1b were sequenced exclusively by this method. All 73 negative control samples interspersed



- 67 -

° among the test samples were negative for HCV RNA.

The amplified DNA fragment obtained in 50 of the 52 HCV isolates was specifically designed to overlap with previously obtained 5'NC sequences (Bukh et al. (1992b) Proc. Natl. Acad. Sci. U.S.A. 89:4942-4946) and with the E1 sequences disclosed herein at approximately 80 nucleotide positions each. A complete match was observed in 6033 of 6035 overlapping nucleotides. Two discrepancies were observed in isolate US6 at nt 552 (C and T) and nt 561 (C and T) respectively. This may have been due to microheterogeneity at these nucleotide positions, since the remaining overlapping sequence was unique for isolate US6. In addition, there were 3 confirmed instances of microheterogeneity: nt 33 in isolate SA11 (C,T and T), nt 36 in isolate S45 (A,C and A), and nt 552 in isolate P10 (C,T and T). Overall, the excellent agreement in these overlapping sequences in this study with the NC sequences disclosed in Bukh et al. and with the E1 sequences disclosed herein definitively ruled out contamination as a source of non-authentic HCV sequences. Furthermore, this analysis proved that the sequences obtained were from a single population, and not from different populations as could happen in mixed infections.

The core (C) gene was exactly 573 nucleotides in length in all 52 HCV isolates with an amino terminal start codon and no in-frame stop codons. Microheterogeneity was observed in 26 of the 52 HCV isolates at 0.2-1.4% of the 573 nucleotide positions of the C gene, and resulted in changes in 0.5-1.0% of the 191 predicted amino acids in 12 of these isolates. A multiple sequence alignment was performed and it showed that the nucleotide identities of the C gene among these HCV isolates were in the range of 79.4-99.0%. In order to compare the genetic relatedness of HCV isolates in different gene regions, phylogenetic trees of the C gene of all 52 HCV isolates and the E1 gene of 51 HCV isolates were constructed using the unweighted pair-

- 68 -

group method with arithmetic mean (Nei, M. (1987) Molecular Evolutionary Genetics (Columbia University Press, New York, N.Y., pp. 287-326) (Figure 8). In both dendrograms a division of the 45 HCV isolates from which C and E1 genes had been cloned into at least six major genetic groups (genotypes 1-6) and 12 minor genetic groups (genotypes I/1a, II/1b, III/2a, IV/2b, 2c, V/3a, 4a-4d, 5a, and 6a) was observed. It is noteworthy that a major division in genetic distance between HCV isolates of genotype 2 and those of the other genotypes in the phylogenetic analyses of both gene sequences was observed. Furthermore, the divergence of the minor genotypes within genotype 2 exhibited a degree of heterogeneity that is equivalent to that observed among the major genotypes. Analysis of the C gene from isolates Z5 and Z8, which had a unique 5' NC sequence (Bukh et al. (1992)) but from which the E1 gene could not be amplified, revealed that these isolates represented two additional genotypes. The designations 4e and 4f are assigned to these genotypes that have not been described previously. Overall, the present specification demonstrates that the genetic relatedness of HCV isolates is equivalent when analyzing the most conserved gene (C) and one of the most variable genes (E1) of the HCV genome, thereby providing strong evidence for the suggested division into major and minor genotypes.

#### Example 4

##### Computer Analysis of the Nucleotide and Deduced Amino Acid Sequences Of The Core Gene Of 52 HCV Isolates

In order to study further the heterogeneity of the C gene, a consensus sequence of the core gene from the 52 HCV isolates (Fig. 6J) was obtained. A total of 335 (58.5%) of the 573 nucleotides of the C gene were invariant among these HCV isolates. Nucleotides at the 1st and 2nd codon positions were invariant at 70.7% and 81.7% of these positions, respectively, while nucleotides at the 3rd

- 69 -

- ° position were invariant at only 23.0% of such positions. Stretches of 6 or more invariant nucleotides were observed from nucleotides 1-8, 22-27, 85-92, 110-125, 131-141, 334-340, 364-371, 397-404, and 511-516 and may be suitable for anchoring primers for amplification of HCV RNA in cDNA PCR assays.

5 Genotype-specific nucleotide positions of the core gene of hepatitis C virus were also noted for each of the genotypes. These genotype-specific nucleotides are shown below where each genotype-specific nucleotide is given in parentheses next to the nucleotide position in which it is found.

10 Genotype 1: 460 (C), 466 (C), 483 (C), 486 (G).

15 Genotype I/1a: 180 (T).

Genotype II/1b: 106 (C), 273 (G).

Genotype 2: 192 (C), 201 (A), 203 (A), 207 (G), 210 (C), 221 (A), 231 (A), 232 (A), 341(A).

20 Genotype III/2a: 315 (C), 355 (G).

Genotype IV/2b: 45 (A), 174 (G), 216 (C), 348 (A), 376 (A), 414 (T).

25 Genotype 2c: 233 (G), 312 (C), 318 (A), 456 (C), 462 (G), 543 (C), 556 (T).

30 Genotype V/3a: 47 (T), 84 (A), 106 (G), 126 (A), 150 (T), 212 (G), 216 (A), 300 (A), 491 (T), 559 (C), 560 (A), 568 (G), 571 (A), 572 (G).

Genotype 4: 59 (T).

35 Genotype 4a: 213 (A), 231 (G), 415 (A).

- 70 -

° Genotype 4b: 66 (G), 145 (G), 310 (A).

Genotype 4c: 213 (T), 219 (A), 270 (T).

Genotype 4d: 212 (T), 327 (G), 469 (C).

5

Genotype 4e: 199 (C), 306 (A), 326 (A).

Genotype 4f: 57 (T), 75 (A), 267 (A).

10 Genotype 5a: 291 (G), 294 (C).

Genotype 6a: 59 (C), 175 (A), 195 (A), 198 (A), 214 (C),  
224 (A), 316 (C), 351 (G), 387 (G), 444-447 (GGCT), 450  
(G), 471-472 (AA), 474 (C).

15

These genotype-specific nucleotides are of utility in designing the genotype-specific PCR primers and hybridization probes.

20

Finally, although the full length nucleic acid sequence of the C gene of isolates representing genotypes I/1a, II/1b, III/2a, IV/2b and V/3a have been reported by others, those of 9 of the 14 genotypes (i.e., 2c, 4a-4f, 5a and 6a) have not been reported previously. In sum, by aligning the consensus sequences of the major genotypes, the present application enables those skilled in the art to map universally conserved sequences as well as genotype-specific sequences of the C gene among 14 genotypes of HCV.

25

In order to study the heterogeneity of the deduced C protein, a multiple sequence alignment of the predicted amino acids for all 52 HCV isolates was performed, and a consensus sequence was obtained (Fig. 7J). The identities of the predicted 191 amino acids of the C protein among these HCV isolates were in the range of 85.3-100.0%. A total of 132 (69.1%) of the 191 amino acids of the C protein were invariant. The most prevalent amino acids in the consensus sequence were glycine (13.6%),

30

35

- 71 -

° arginine (12.6%), proline (11.0%), and leucine (9.9%). The most conserved amino acids were tryptophan (5 of 5 amino acids invariant), aspartic acid (5 of 5 amino acids invariant), proline (19 of 21 amino acids invariant) and glycine (23 of 26 amino acids invariant). Previous analyses indicated that HCV is evolutionarily related to pestiviruses (Miller et al. (1990) Proc. Natl. Acad. Sci. U.S.A. 87:2057-2061). In this regard, it is of interest to note that the C proteins of both viruses have a high content of proline residues (Collette M.S. et al. (1988) Virology 165:200-208), which are likely to be important in maintaining the structure of this protein. As is characteristic for a protein that binds to nucleic acid, the C protein has conserved amino acids that are basic and positively charged, and these are capable of neutralizing the negative charge of the HCV RNA encapsidated by this protein (Rice, C.M. et al. (1986) in *Togaviridae and Flaviviridae*, eds Schleinger, S. & Schlensinger, M.J. (Plenum Press, New York, N.Y.) pp. 279-326). Specifically, over 16% of the amino acids in the consensus sequence of the C protein of HCV are arginine and lysine that are located primarily in three clusters (i.e., from amino acids 6-23, 39-74 and 101-121) (Shih, C.M. et al. (1993) J. Gen. Virol. 67:5823-5832) (Fig. 7J). The 10 arginine and lysine residues within amino acids 39-62 are invariant among all 52 HCV isolates, suggesting that this domain may represent an important RNA-binding site. The capsid proteins of the related flavi- and pestiviruses (Miller et al. (1990)) also have a high content of arginine and lysine (Rice et al. (1986); Collette et al. (1988)). Although there are three major hydrophilic regions (i.e., amino acids 2-23, 39-74 and 101-121) that are conserved in all 52 HCV isolates, the remainder of the C protein is hydrophobic. Interestingly, one such highly conserved hydrophobic domain from aa 24-39 is flanked by proline residues. The hydrophobic domains are likely to be

- 72 -

° involved in protein-protein and/or protein-RNA interactions during assembly of the nucleocapsid, as well as in interaction with the lipoprotein envelope, as has been suggested for flaviviruses (Rice et al. (1986)). Other significant observations are: (i) a cluster of 5 invariant  
5 tryptophan residues from aa 76-107; (ii) the lack of an N-linked glycosylation site (N-X-T/S); (iii) two potential nuclear localization signals (i.e., PRRGPR at amino acids 38-43 and PRGRRQP at amino acids 58-64) that are present in all 52 HCV isolates (Shih et al. (1993)); and (iv) a  
10 putative DNA-binding motif SPRG at amino acids 99-102, found in 51 of the 52 HCV isolates, with SP present in all 52 isolates. This study demonstrates that the C protein has features that are highly conserved among the various genotypes of HCV, and that are known to be characteristic  
15 of capsid proteins of other related viruses.

It should also be noted that the phylogenetic analysis of the amino acid sequence of the C proteins was not capable of resolving the minor groups within genotypes 1 and 4 because of the conservation of this protein (data  
20 not shown). Indeed, only a few type-specific amino acids were identified. One striking example was that isolates of genotype 4 have an additional methionine at position 20 that is specific for this major genetic group. Finally, the conservation of the sequences surrounding the cleavage  
25 site between the C and the E1 proteins of the different genotypes, which has been determined to be between amino acid 191 (alanine) and aa 192 (tyrosine) in HCV isolates of genotype 1 was analyzed (Hijikata, M., et al. (1991) Proc. Natl. Acad. Sci. USA 88:5547-5551). The C-terminal  
30 sequence of C is serine-alanine in all but one of the 48 HCV isolates comprising genotypes 1, 2, 4, 5 and 6. However, all 4 HCV isolates of genotype 3 in this study, as well as isolates of genotype 3 published previously (Okamoto, H., et al. (1993) *J. Gen. Virol.* 74:2385-2390,  
35 Stuyver, L., et al. (1993) *Biochem. Biophys. Res. Comm.*

- 73 -

- ° 192:635-641), contain alanine-serine at this position. Thus, studies will be needed to determine the C/E1 cleavage site in genotype 3 isolates. Overall, the present invention application discloses the mapping of universally conserved sequences, as well as genotype-specific sequences, of the C protein among 14 genotypes of HCV.

**Implications of the mapping of universally conserved and genotype-specific core nucleotide and amino acid core sequences for diagnosis of HCV infection and for determination of HCV genotypes**

- 10 Detection of antibodies directed against the HCV core protein is important in the diagnosis of HCV infection. The recombinant C22-3 protein, spanning amino acids 2-120 of the C gene, is a major component of the commercially available second-generation anti-HCV tests.
- 15 Several studies have indicated that the three major hydrophilic regions of the C protein contain linear immunogenic epitopes (summarized in J. Clin. Microbiol., 30:1989-1994) (Sällberg, M. et al. (1992)). For example, antibodies against synthetic peptides from amino acids 1-18, 51-68 and 101-118 were detected in infected patients (Sällberg, M. et al. (1992)). The present application demonstrates that, while these immunogenic regions are highly conserved, genotype-specific differences are observed at several amino acid positions that may influence the specificity and sensitivity of the serological tests.
- 25 One such example is that a single amino acid substitution at amino acid 110 has been demonstrated to affect seroreactivity (Sällberg, et al. (1992)). Despite the high degree of conservation in the immunodominant regions of the C protein among the different genotypes, it is possible that genetic heterogeneity of the C protein could lead to false negative results in current serological tests.

- 30 With respect to genotype analysis, several methods have been used to determine the genotype of HCV isolates without resorting to sequence analysis. These
- 35

- 74 -

° include PCR followed by: (i) amplification with type-specific primers (Okamoto, H. et al. (1992) J. Gen. Virol., 73:673-679); (ii) determination of restriction-length polymorphism (Simmons, P. et al. (1993) J. Gen. Virol., 74:661-668); and (iii) specific hybridization (Stuyver, L. 5 (1993) J. Gen. Virol., 74:1093-1102). The proposed methods have primarily been based on 5' NC and C sequences. Previous studies suggested that 5' NC-based genotyping systems would only be predictive of the major genetic groups of HCV (Bukh, J., et al. (1992) Proc. Natl. Acad. Sci. USA 89:4942- 4946, Bukh, J., et al. (1993) Proc. Natl. Acad. Sci. USA 90:8234-8238). The most widely used C-based genotype system has been the PCR assay with type-specific primers that was designed for distinguishing HCV isolates of genotypes I/1a, II/1b, III/2a, IV/2b and V/3a 15 (Okamoto, H., et al. (1993) J. Gen. Virol. 74:2385-2390, Okamoto, H. et al. (1992) J. Gen. Virol. 73:673-679). Since this system was developed prior to the identification of genotypes 2c, 4a-4f, 5a and 6a there are significant limitations to this typing system. For example, the 20 primers specific for genotype IV/2b (nt 270-251) are as highly conserved within isolates of genotype 4c and 6a as within the isolates of genotype IV/2b. Thus, this assay probably can not distinguish among these genotypes. Another C-based approach involves distinguishing between genotypes 25 1 and 2 by type-specific antibody responses (Machida et al (1992) Hepatology, 16:886-891). Synthetic peptides composed of amino acids 65-81 were found to be genotype-specific for genotypes 1 and 2 in ELISA assays. The present analysis of amino acid sequences demonstrated 30 significant variation within isolates of genotypes 1 and 2. Thus it is likely that these peptides will not identify all isolates of genotypes 1 and 2. Furthermore, the peptide for genotype 1 was highly conserved within isolates of genotypes 3 and 4 and might detect antibodies against these 35 genotypes as well. Finally, it should be pointed out that



- 75 -

- ° most isolates of genotypes 3 and 4 had an identical amino acid sequence at positions 65-81.

#### Example 5

##### 5 Detection by ELISA Based on Antigen from Insect Cells Expressing Complete E1 Or Core Protein

Expression of E1 or Core protein in SF9 cells. A cDNA (eg SEQ ID NO:1) encoding a complete E1 protein (eg SEQ ID NO:52) or a cDNA (eg SEQ ID NO:103) encoding a complete core protein (e.g. SEQ ID NO:155) is subcloned into pBlueBac - Transfer vector (Invitrogen) using standard subcloning procedures. The resultant recombinant expression vector is cotransfected into SF9 insect cells (Invitrogen) by the Ca precipitation method according to the Invitrogen protocol.

15 ELISA Based on Infected SF9 cells.  $5 \times 10^6$  SF9 cells infected with the above-described recombinant expression vector are resuspended in 1 ml of 10 mM Tris-HCl, pH 7.5, 0.15M NaCl and are then frozen and thawed 3 times. 10 ul of this suspension is dissolved in 10 ml of carbonate buffer (pH 9.6) and used to cover one flexible microtiter assay plate (Falcon). Serum samples are diluted 20 1:20, 1:400 and 1:8000, or 1:100, 1:1000 and 1:10000. Blocking and washing solutions for use in the ELISA assay are PBS containing 10% fetal calf serum and 0.5% gelatin (blocking solution) and PBS with 0.05% Tween -20 (Sigma, 25 St.Louis, MO) (washing solution). As a secondary antibody, peroxidase-conjugated goat IgG fraction to human IgG or horse radish peroxidase-labelled goat anti-Old or anti-New World monkey immunoglobulin is used. The results are 30 determined by measuring the optical density (O.D.) at 405 nm.

To determine if insect cells-derived E1 or core protein representing genotype I/a of HCV could detect anti-HCV antibody in chimpanzees infected with genotype I/1a of HCV, three infected chimpanzees are examined. The serum of 35

- 76 -

- ° all 3 chimpanzees are found to seroconvert to anti-HCV.

#### Example 6

##### Use of the Complete E1 Protein as a Vaccine

5 Mammals are immunized with purified or partially purified E1 protein in an amount sufficient to stimulate the production of protective antibodies. The immunized mammals challenged with various genotypes of HCV are protected.

10 It is understood by one skilled in the art that the recombinant E1 protein used in the above vaccine can also be used in combination with other recombinant E1 proteins having an amino acid sequence shown in SEQ ID NOS:52-102. In addition, recombinant core proteins having  
15 an amino acid sequence shown in SEQ ID NOS:155-206 could also be used in the above vaccine, either alone, in combination with other recombinant core proteins of the present invention, or in combination with recombinant E1 proteins having an amino acid sequence shown in SEQ ID  
20 NOS:52-102.

#### Example 7

##### Determination of the Genotype of an HCV Isolate Via Hybridization of Genotype-Specific Oligonucleotides to RT-PCR Amplification Products.

25 Viral RNA is isolated from serum obtained from a mammal and is subjected to RT-PCR as in Example 1 or Example 3. Following amplification, the amplified DNA is purified as described in Example 1 or Example 3 and  
30 aliquots of 100 ul of amplification product are applied to dots on a nitrocellulose filter set in a dot blot apparatus. The dots are then cut into separate dots and each dot is hybridized to a <sup>32</sup>P-labelled oligonucleotide specific for a single genotype of HCV. The  
35 oligonucleotides to be used as hybridization probes are deduced from the consensus sequences shown in Figures 1A-1H

- 77 -

- ° or 6A-6J or from the SEQ ID NOs: representing E1 or core sequences comprising genotypes 4a-4f, 2c and 6a.

#### Example 8

##### ELISA Based on Synthetic Peptides Derived From E1 cDNA Sequences

5 E1 peptide(s) specific for genotype I/1a is placed in 0.1% PBS buffer and 50ul of a 1mg/ml solution of peptide is used to cover each well of the microtiter assay plate. Serum samples from two mammals infected with  
10 genotype I/1a HCV and from one mammal infected with genotype 5a HCV are diluted as in Example 3 and the ELISA is carried out as in Example 3. Both mammals infected with genotype I HCV react positively with peptides while the  
15 mammal infected with genotype 5a HCV exhibits no reactivity. One skilled in the art would readily understand that in the above experiment, core peptides specific for genotype I/1a could be used in place of, or in combination with the E1 genotype-specific peptide(s).

#### Example 9

##### Use of E1 Peptides as a Vaccine

20 Since the E1 genotype-specific peptides of the present invention are derived from two variable regions in the complete E1 protein, there exists support for the use  
25 of these peptides as a vaccine to protect against a variety of HCV genotypes. Mammals are immunized with peptide(s) selected from SEQ ID NOs: 136-159 in an amount sufficient to stimulate production of protective antibodies. The  
30 immunized mammals challenged with various genotypes of HCV are protected. One skilled in the art would readily understand that genotype-specific core peptides of the present invention could also be used either alone, in combination with each other, or in combination with the  
35 genotype-specific E1 peptides, as a vaccine to protect against a variety of HCV genotypes. In addition, the above

- 78 -

- ° vaccines may also be formulated using the universal core and/or E1 peptides of the present invention.

5

10

15

20

25

30

35

- 79 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANTS: BUKH, J., MILLER, R.H. AND  
PURCELL, R.H.

(ii) TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED  
AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND  
CORE GENES OF ISOLATES OF HEPATITIS C VIRUS  
AND THE USE OF REAGENTS DERIVED FROM THESE  
SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

(iii) NUMBER OF SEQUENCES: 263

(iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: MORGAN & FINNEGAN  
(B) STREET: 345 PARK AVENUE  
(C) CITY: NEW YORK  
(D) STATE: NEW YORK  
(E) COUNTRY: USA  
(F) ZIP: 10154

(v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: FLOPPY DISK  
(B) COMPUTER: IBM PC COMPATIBLE  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: WORDPERFECT 5.1

(vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER: TO BE ASSIGNED  
(B) FILING DATE: 15-AUG-1995  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NO: 08/086,428  
(B) FILING DATE: 29 JUNE 1993

(vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: 08/290/665  
(B) FILING DATE: 15 AUGUST 1994

(viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: RICHARD W. BORK  
(B) REGISTRATION NUMBER: 36,459  
(C) REFERENCE/DOCKET NUMBER: 2026-4116

(viii) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (212) 758-4800  
(B) TELEFAX: (212) 751-6849  
(C) TELEX: 421792

(2) INFORMATION FOR SEQ ID NO:1:

- 80 -

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK7

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

10	TAC	CAA	GTG	CGC	AAC	TCC	ACG	GGG	CTT	TAC	CAT	GTC	ACC	39
	AAT	GAT	TGC	CCT	AAC	TCG	AGT	ATC	GTG	TAC	GAG	GCG	GCC	78
	GAT	GCC	ATC	CTG	CAC	ACT	CCG	GGG	TGT	GTC	CCT	TGC	GTT	117
	CGC	GAG	GGT	AAC	GTC	TCG	AGG	TGT	TGG	GTG	GCG	ATG	ACC	156
	CCC	ACG	GTG	GCC	ACC	AGG	GAT	GGC	AAA	CTC	CCC	ACA	GCG	195
	CAG	CTT	CGA	CGT	CAC	ATC	GAT	CTG	CTC	GTC	GGG	AGT	GCC	234
	ACC	CTC	TGT	TCG	GCC	CTC	TAC	GTG	GGG	GAC	CTG	TGC	GGG	273
	TCT	GTC	TTT	CTT	GTC	GGT	CAA	CTG	TTT	ACC	TTC	TCT	CCC	312
15	AGG	CGC	CAC	TGG	ACG	ACG	CAA	GGC	TGC	AAT	TGT	TCT	ATC	351
	TAT	CCT	GGC	CAT	ATA	ACG	GGT	CAC	CGC	ATG	GCG	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCC	CCT	ACC	ACG	GCG	TTG	GTA	GTA	429
	GCT	CAG	CTG	CTC	CGG	ATC	CCG	CAA	GCC	ATC	TTG	GAC	ATG	468
	ATC	GCT	GGT	GCT	CAC	TGG	GGA	GTC	CTG	GCG	GGC	ATA	GCG	507
	TAT	TTT	TCC	ATG	GTG	GGG	AAC	TGG	GCG	AAG	GTC	CTG	GTA	546
	GTG	CTG	CTG	CTA	TTT	GCC	GGC	GTC	GAC	GCG				576

- (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK9

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

30	TAC	CAA	GTA	CGC	AAC	TCC	TCG	GGC	CTC	TAC	CAT	GTC	ACC	39
	AAT	GAT	TGC	CCT	AAC	TCG	AGT	ATT	GTG	TAC	GAG	GCG	GCC	78
	GAT	GCC	ATC	CTG	CAT	TCT	CCA	GGG	TGT	GTC	CCT	TGC	GTT	117
	CGC	GAG	GGT	AAC	GCC	TCG	AAA	TGT	TGG	GTG	GCG	GTG	GCC	156
	CCC	ACG	GTG	GCC	ACC	AGG	GAC	GGC	AAG	CTC	CCC	GCA	ACG	195
	CAG	CTT	CGA	CGT	CAC	ATC	GAT	CTG	CTT	GTC	GGG	AGC	GCC	234
	ACC	CTC	TGC	TCG	GCC	CTC	TAT	GTG	GGG	GAC	TTG	TGC	GGG	273
35	TCT	GTC	TTC	CTT	GTC	GGC	CAA	CTG	TTC	ACC	TTC	TCC	CCC	312
	AGA	CGC	CAC	TGG	ACA	ACG	CAA	GAC	TGC	AAC	TGT	TCT	ATC	351

- 81 -

TAC CCC GGC CAT ATT ACG GGT CAT CGC ATG GCG TGG GAT 390  
 ATG ATG ATG AAC TGG TCC CCT ACA GCA GCG CTG GTA ATG 429  
 GCG CAG CTG CTC AGG ATC CCG CAG GCC ATC TTG GAC ATG 468  
 ATC GCT GGT GCC CAC TGG GGA GTC CTA GCG GGC ATA GCG 507  
 TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC GTG GTG 546  
 GTA CTG TTG CTG TTT ACC GGC GTC GAT GCG 576

5

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DR1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

15

CAC CAA GTG CGC AAC TCT ACA GGG CTT TAC CAT GTC ACC 39  
 AAT GAT TGC CCT AAT TCG AGT ATT GTG TAC GAG GCG GCC 78  
 GAT GCC ATC CTG CAC GCG CCG GGG TGT GTC CCT TGC GTT 117  
 CGC GAG GGT AAC GCC TCG AGG TGT TGG GTG GCG GTG ACC 156  
 CCC ACG GTG GCC ACC AGG GAC GGC AAA CTC CCC ACA ACG 195  
 CAG CTT CGA CGT CAC ATC GAC CTG CTT GTC GGG AGC GCC 234  
 ACC CTC TGC TCG GCC CTC TAC GTG GGG GAC CTG TGC GGG 273  
 TCT GTC TTC CTT GTC GGT CAA CTG TTC ACC TTT TCT CCC 312  
 AGG CGC CAC TGG ACA ACG CAA GAC TGC AAT TGT TCT ATC 351  
 TAT CCC GGC CAT ATA ACG GGA CAC CGT ATG GCA TGG GAT 390  
 ATG ATG ATG AAC TGG TCC CCT ACG ACA GCG CTG GTA ATG 429  
 GCT CAG CTG CTC CGG ATC CCA CAA GCC ATC TTG GAC ATG 468  
 ATC GCT GGA GCC CAC TGG GGA GTC CTA GCG GGC ATA GCG 507  
 TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC GTG GTA 546  
 GTG CTG TTG CTG TTT GCC GGC GTT GAT GCG 576

20

25

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DR4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

35

- 82 -

CAC CAA GTG CGC AAC TCT ACA GGG CTT TAC CAT GTC ACC 39  
 AAT GAT TGC CCT AAT TCG AGT ATT GTG TAC GAG GCG GCC 78  
 GAT GCC ATC CTG CAC ACG CCG GGG TGT GTC CCT TGC GTT 117  
 CGC GAG GGT AAC ACC TCG AGG TGT TGG GTG GCG GTG ACC 156  
 CCC ACG GTG GCC ACC AGG GAC GGC AAA CTC CCC ACA ACG 195  
 CAG CTC CGA CGT CAC ATC GAC CTG CTT GTC GGG AGC GCC 234  
 5 ACC CTC TGC TCG GCC CTC TAC GTG GGG GAC TTG TGC GGG 273  
 TCT GTC TTC CTT GTC GGT CAA CTG TTC ACC TTC TCT CCC 312  
 AGG CAC CAC TGG ACA ACG CAA GAC TGC AAT TGT TCC ATC 351  
 TAT CCC GGC CAT ATA ACG GGC CAC CGC ATG GCG TGG GAT 390  
 ATG ATG ATG AAC TGG TCC CCT ACG ACA GCG CTG GTA GTA 429  
 GCT CAG CTG CTC CGG ATC CCA CAA GCC ATC TTG GAC ATG 468  
 ATC GCT GGT GCC CAC TGG GGA GTC CTA GCG GGC ATA GCG 507  
 10 TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC CTG GTA 546  
 GTG CTG TTG CTG TTT GCC GGC GTT GAT GCG 576

## (2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 15 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S14

## 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TAC CAA GTG CGC AAC TCC ACG GGG CTT TAC CAT GTT ACC 39  
 AAT GAT TGC CCT AAC TCG AGT ATT GTG TAC GAG ACA GCT 78  
 GAT GCT ATC CTA CAC GCT CCG GGA TGT GTC CCT TGC GTT 117  
 CGT GAG GGT AAC ACC TCG AGG TGT TGG GTG GCG ATG ACC 156  
 CCC ACG GTG GCC ACC AGG GAC GGC AAA CTC CCC GCA ACG 195  
 25 CAG CTT CGA CGT TAC ATC GAT CTG CTT GTC GGG AGC GCC 234  
 ACC CTC TGT TCG GCC CTC TAC GTG GGG GAC TTG TGC GGG 273  
 TCT GTC TTT CTT GTC GGT CAG CTG TTT ACC TTC TCT CCC 312  
 AGG CGC CTC TGG ACG ACG CAA GAC TGC AAT TGT TCT ATC 351  
 TAT CCC GGC CAT ATA ACG GGT CAT CGC ATG GCA TGG GAT 390  
 ATG ATG ATG AAC TGG TCC CCT ACG ACG GCA CTG GTA GTA 429  
 GCT CAG CTG CTC CGG ATC CCA CAA GCC ATC TTG GAT ATG 468  
 ATC GCT GGT GCT CAC TGG GGA GTC CTA GCG GGC ATA GCG 507  
 30 TAT TTC TCC ATG GTG GGA AAC TGG GCG AAG GTC CTA GTG 546  
 GTG CTG CTG CTA TTC GCC GGC GTT GAC GCG 576

## (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid



- 83 -

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: S18

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

	TAC	CAA	GTA	CGC	AAC	TCC	ACG	GGC	CTT	TAC	CAT	GTC	ACC	39
	AAT	GAC	TGC	CCT	AAC	TCG	AGC	ATT	GTG	TAC	GAG	ACG	GCC	78
	GAT	ACC	ATC	CTA	CAC	TCT	CCG	GGG	TGT	GTC	CCT	TGC	GTT	117
	CGC	GAG	GGT	AAC	GCC	TCG	AGA	TGT	TGG	GTG	CCG	GTG	GCC	156
	CCC	ACA	GTT	GCC	ACC	AGG	GAC	GGC	AAA	CTC	CCC	GCA	ACG	195
10	CAG	CTT	CGA	CGT	CAC	ATC	GAT	CTG	CTT	GTT	GGG	AGC	GCC	234
	ACC	CTC	TGC	TCG	GCC	CTC	TAT	GTG	GGG	GAC	CTG	TGC	GGG	273
	TCT	GTC	TTT	CTT	GTC	AGC	CAG	CTG	TTC	ACT	ATC	TCC	CCC	312
	AGG	CGC	CAC	TGG	ACA	ACG	CAA	GAC	TGC	AAC	TGT	TCT	ATC	351
	TAC	CCC	GGC	CAT	ATA	ACG	GGT	CAC	CGT	ATG	GCA	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCC	CCT	ACA	ACG	GCG	TTG	GTA	ATA	429
	GCT	CAG	CTG	CTC	AGG	GTC	CCG	CAA	GCC	GTC	TTG	GAC	ATG	468
15	ATC	GCT	GGT	GCC	CAC	TGG	GGA	GTC	CTA	GCG	GGC	ATA	GCG	507
	TAT	TTC	TCC	ATG	GCG	GGG	AAC	TGG	GCG	AAG	GTC	CTG	CTA	546
	GTG	CTG	TTG	CTG	TTT	GCC	GGC	GTC	GAT	GCG				576

(2) INFORMATION FOR SEQ ID NO:7:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 576 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
25 (C) INDIVIDUAL ISOLATE: SW1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

	TAC	CAA	GTA	CGC	AAC	TCC	TCG	GGC	CTT	TAC	CAT	GTC	ACC	39
	AAT	GAT	TGC	CCT	AAC	TCG	AGT	ATT	GTG	TAC	GAG	ACG	GCC	78
	GAT	GCC	ATT	CTA	CAC	TCT	CCA	GGG	TGT	GTC	CCT	TGC	GTT	117
	CGC	GAG	GAT	GGC	GCC	CCG	AAG	TGT	TGG	GTG	GCG	GTG	GCC	156
30	CCC	ACA	GTC	GCC	ACT	AGG	GAC	GGC	AAA	CTC	CCT	GCA	ACG	195
	CAG	CTT	CGA	CGT	CAC	ATC	GAT	CTG	CTT	GTC	GGA	AGC	GCC	234
	ACC	CTC	TGC	TCG	GCC	CTC	TAC	GTG	GGG	GAC	TTG	TGC	GGG	273
	TCT	GTC	TTT	CTC	GTC	AGT	CAA	CTG	TTC	ACG	TTC	TCC	CCC	312
	AGG	CGC	CAC	TGG	ACA	ACG	GGT	CAC	GCG	ATG	GCA	TGG	GAT	351
	TAT	CCC	GGC	CAC	ATA	ACG	GGT	CAC	GCG	ATG	GCA	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCC	CCC	ACA	ACA	GCG	CTG	GTA	GTA	429
35	GCT	CAG	CTG	CTC	AGG	ATC	CCG	CAA	GCC	GTC	TTG	GAC	ATG	468
	ATC	GCT	GGT	GCC	CAC	TGG	GGA	GTC	CTA	GCG	GGC	ATA	GCG	507

- 84 -

TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC CTG ATA 546  
 GTG CTG TTG CTG TTT TCC GGC GTC GAT GCG 576

## (2) INFORMATION FOR SEQ ID NO:8:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- (vi) ORIGINAL SOURCE:  
 10 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US11

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

	TAC	CAA	GTA	CGC	AAC	TCC	ACG	GGG	CTT	TAC	CAT	GTC	ACC	39
	AAT	GAT	TGC	CCT	AAC	TCG	AGT	ATT	GTG	TAC	GAG	GCG	GCC	78
	GAT	GCC	ATC	CTG	CAC	ACT	CCG	GGG	TGT	GTT	CCT	TGC	GTT	117
15	CGC	GAG	GGT	AAC	GCT	TCG	AGG	TGT	TGG	GTG	GCG	ATG	ACC	156
	CCC	ACG	GTG	GCC	ACC	AGG	GAC	GGC	AAA	CTC	CCC	ACA	ACG	195
	CAA	CTT	CGA	CGT	CAC	ATC	GAT	CTG	CTT	GTC	GGG	AGC	GCC	234
	ACC	CTC	TGT	TCG	GCC	CTC	TAC	GTG	GGG	GAC	CTG	TGC	GGG	273
	TCT	GTC	TTT	CTT	GTC	GGT	CAA	CTG	TTT	ACC	TTC	TCT	CCC	312
	AGA	CGC	CAC	TGG	ACG	ACG	CAG	GGC	TGC	AAT	TGT	TCT	ATC	351
	TAT	CCC	GGC	CAT	ATA	ACG	GGT	CAC	CGC	ATG	GCA	TGG	GAT	390
20	ATG	ATG	ATG	AAC	TGG	TCC	CCT	ACG	GCG	GCG	TTG	GTG	GTA	429
	GCT	CAG	CTG	CTC	CGG	ATC	CCA	CAA	GCC	ATC	TTG	GAC	ATG	468
	ATC	GCT	GGT	GCT	CAC	TGG	GGA	GTC	CTA	GCG	GGC	ATA	ACG	507
	TAT	TTC	TCC	ATG	GTG	GGG	AAC	TGG	GCG	AAG	GTC	CTG	GTA	546
	GTG	CTG	CTG	CTA	TTT	GCC	GGC	GTC	GAC	GCG				576

## (2) INFORMATION FOR SEQ ID NO:9:

- 25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- (vi) ORIGINAL SOURCE:  
 30 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: D1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

	TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	GTG	TAC	CAT	GTC	ACG	39
	AAC	GAC	TGT	TCC	AAC	TCG	AGC	ATT	GTG	TAT	GAG	ACA	GCG	78
	GAC	ATG	ATC	ATG	CAC	ACC	CCC	GGG	TGC	GTG	CCC	TGC	GTT	117
35	CGG	GAG	GAC	AAC	TCC	TCT	CGC	TGC	TGG	GTA	GCG	CTC	ACC	156
	CCC	ACG	CTC	GCG	GCT	AGG	AAT	GGC	AAC	GTC	CCC	ACT	ACG	195

- 85 -

GCG ATA CGA CGC CAC GTC GAT TTG CTC GTT GGG GCG GCT 234  
 GCT TTC TGC TCC GCC ATG TAC GTG GGG GAT CTC TGC GGA 273  
 TCT GTT TTC CTC ATC TCC CAG CTG TTC ACC CTC TCG CCT 312  
 CGC CGG CAT GAG ACG GTA CAG GAG TGT AAT TGC TCA ATC 351  
 TAT CCC GGC CAC GTG ACA GGT CAC CGT ATG GCT TGG GAT 390  
 ATG ATG ATG AAC TGG TCA CCT ACA ACA GCC TTA GTG GTA 429  
 5 TCG CAG TTA CTC CGG ATC CCA CAA GCT GTC ATG GAC ATG 468  
 GTG GCG GGG GCC CAC TGG GGG GTC CTG GCG GGC CTC GCC 507  
 TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTC TTT GCT GGC GTT GAC GGC 576

## (2) INFORMATION FOR SEQ ID NO:10:

10

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 576 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

15

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: D3

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TAT GAA GTG CGC AAC GTG TCC GGG GTG TAC CAA GTC ACC 39  
 AAT GAC TGT TCC AAC TCG AGC ATC GTG TAT GAG ACA GCG 78  
 GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC GTT 117  
 20 CGG GAG GAC AAC TCC TCT CGC TGC TGG GTA GCG CTC ACC 156  
 CCC ACG CTC GCG GCT AGG AAT AGC AGC GTC CCC ACT ACG 195  
 ACA ATA CGA CGC CAC GTC GAT TTG CTC GTT GGG GCG GCT 234  
 GCT TTC TGC TCC GCC ATG TAC GTG GGG GAT CTT TGC GGA 273  
 TCT GTT TTC CTC GTC TCC CAG CTG TTC ACC TTC TCG CCT 312  
 CGC CGG CAT GAG ACA GTA CAG GAA TGT AAC TGC TCA ATC 351  
 TAT CCC GGC CAC GTG ACA GGT CAC CGC ATG GCT TGG GAT 390  
 ATG ATG ATG AAC TGG TCG CCT ACA GCA GCC CTA GTG GTA 429  
 25 TCG CAG TTA CTC CGG ATC CCA CAA GCT GTC GTG GAC ATG 468  
 GTG GCG GGG GCC CAC TGG GGG GTC CTG GCG GGC CTC GCC 507  
 TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTC TTT GCT GGC GTC GAC GGC 576

## (2) INFORMATION FOR SEQ ID NO:11:

30

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 576 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

35

- (A) ORGANISM: homosapiens

- 86 -

## (C) INDIVIDUAL ISOLATE: DK1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

	TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	GTG	TAC	CAC	GTC	ACA	39
	AAC	GAC	TGC	TCC	AAC	TCA	AGC	ATC	GTG	TAT	GAG	GCA	GTG	78
5	GAC	GTG	ATC	ATG	CAT	ACC	CCA	GGG	TGC	GTG	CCC	TGC	GTT	117
	CGG	GAG	AAC	AAC	CAC	TCC	CGT	TGC	TGG	GTA	GCG	CTC	ACC	156
	CCC	ACG	CTC	GCG	GCC	AGG	AAC	GCC	AGC	ATC	CCC	ACT	ACG	195
	ACA	ATA	CGA	CGC	CAT	GTC	GAT	TTG	CTC	GTT	GGG	GCG	GCT	234
	GCT	TTC	TGC	TCC	GCT	ATG	TAC	GTG	GGG	GAC	CTC	TGC	GGA	273
	TCC	GTT	TTC	CTC	GTC	TCT	CAG	CTG	TTC	ACC	TTT	TCA	CCT	312
	CGC	CGG	CAT	GAG	ACA	GCA	CAG	GAC	TGC	AAC	TGC	TCA	ATC	351
10	TAT	CCC	GGC	CAC	GTT	TCA	GGT	CAC	CGC	ATG	GCT	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCA	CCT	ACA	ACA	GCC	CTA	GTG	CTA	429
	TCG	CAG	TTA	CTC	CGA	ATC	CCA	CAA	GCT	GTC	GTG	GAC	ATG	468
	GTG	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTC	GCC	507
	TAC	TAC	TCC	ATG	GCG	GGG	AAC	TGG	GCC	AAG	GTT	TTA	ATT	546
	GTG	TTG	CTA	CTC	TTT	GCC	GGC	GTT	GAT	GGG				576

## 15 (2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK3

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

	TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	ATA	TAC	CAT	GTC	ACG	39
25	AAC	GAC	TGC	TCC	AAC	TCA	AGC	GTC	GTG	TAT	GAG	ACA	GCA	78
	GAC	ATG	ATC	ATG	CAT	ACC	CCT	GGA	TGC	GTG	CCC	TGC	GTA	117
	CGG	GAG	AAC	AAC	TCC	TCC	CGC	TGT	TGG	GTA	GCG	CTC	ACT	156
	CCC	ACG	CTC	GCG	GCC	AGG	AAC	GTC	AGC	GTC	CCC	ACC	ACG	195
	ACA	ATA	CGA	CGT	CAC	GTC	GAC	TTG	CTC	GTT	GGG	GCG	GCT	234
	GCC	TTC	TGC	TCC	GCT	ATG	TAC	GTG	GGG	GAT	CTC	TGC	GGA	273
	TCT	GTT	TTC	CTT	GTC	TCC	CAG	CTG	TTC	ACC	TTC	TCG	CCT	312
	CGC	CGA	CAC	GAG	ACA	GTA	CAG	GAC	TGC	AAC	TGC	TCA	CTC	351
30	TAT	CCC	GGC	CAC	GTA	TCA	GGT	CAC	CGC	ATG	GCT	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCC	CCT	ACA	GCA	GCC	CTA	GTG	GTG	429
	TCG	CAA	TTA	CTC	CGG	ATC	CCG	CAA	GCT	GTC	GTG	GAC	ATG	468
	GTG	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTA	GCG	GGC	CTT	GCC	507
	TAC	TAT	TCC	ATG	GTG	GGA	AAC	TGG	GCT	AAG	GTT	TTG	ATT	546
	GTG	ATG	CTA	CTT	TTT	GCC	GGC	GTT	GAT	GGG				576

35

- 87 -

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

10	CAT GAA GTG CAC AAC GTA TCC GGG ATC TAC CAT GTC ACG	39
	AAC GAC TGC TCC AAC TCA AGT ATT GTG TAT GAG GCA GCG	78
	GAC ATG ATC ATG CAT ACC CCC GGG TGC GTG CCC TGC GTC	117
	CGG GAG AAC AAC TCC TCC CGT TGC TGG GTA GCG CTC ACT	156
	CCC ACG CTC GCG GCC AGG AAC GCC AGC ATC CCC ACT ACG	195
	ACA ATA CGA CGC CAT GTC GAC TTG CTC GTT GGG GCG GCT	234
	GCT TTC TGC TCC GCC ATG TAC GTG GGA GAT CTC TGC GGA	273
	TCT GTC TTC CTC GTC TCC CAG TTG TTC ACC TTC TCG CCT	312
15	CGC CGG CAT GAG ACG GTA CAG GAC TGC AAT TGC TCA ATC	351
	TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGG GAT	390
	ATG ATG ATG AAC TGG TCA CCT ACA GCA GCC CTA GTG GTA	429
	TCG CAG TTA CTC CGA CTC CCA CAA GCT GTC ATG GAC ATG	468
	GTG GCG GGA GCC CAC TGG GGA GTC CTA GCG GGC CTT GCT	507
	TAC TAT TCC ATG GTG GGG AAC TGG GCC AAG GTT TTG ATT	546
	GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG	576

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

30	TAT GAA GTG CGC AAC GTG TCC GGG GTA TAC CAT GTC ACG	39
	AAC GAC TGC TCC AAC TTA AGC ATC GTG TAC GAG ACA ACG	78
	GAC ATG ATC ATG CAC ACC CCT GGG TGC GTG CCC TGC GTT	117
	CGG GAA AAC AAC TCC TCC CGT TGT TGG GTA GCG CTC GCC	156
	CCC ACG CTC GCG GCC AGG AAC GCC AGC GTC CCC ACC ACG	195
	GCA ATA CGA CGC CAC GTC GAC TTG CTC GTT GGG GCG GCT	234
	GCT TTC TGC TCC GCT ATG TAC GTG GGG GAT CTT TGC GGA	273
35	TCT GTT TTC CTC GTC TCC CAG CTG TTC ACC TTC TCG CCT	312
	CGC CGA CAC GAG ACG GTA CAG GAC TGC AAC TGC TCA ATC	351

- 88 -

TAT CCC GGC CAC GTA ACA GGT CAC CGC ATG GCT TGG GAT 390  
 ATG ATG ATG AAC TGG TCA CCT ACA ACA GCC CTA GTG GTG 429  
 TCG CAG TTA CTC CGG ATC CCG CAA GCT GTC GTG GAC ATG 468  
 GTA GCG GGG GCC CAC TGG GGG GTC CTG GCG GGC CTT GCC 507  
 TAC TAT TCC ATG GTG GGA AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTT TTT GCC GGC GTT GAT GGG 576

5

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

15

TAT GAA GTG CGC AAC GTG TCC GGG ATA TAC CAT GTC ACG 39  
 AAC GAC TGC TCC AAC TCA AGC ATC GTG TAT GAA ACA GCG 78  
 GAC ATG ATT ATG CAT ACC CCT GGA TGC ATG CCC TGC GTT 117  
 CGG GAG AAC AAC TCC TCC CGT TGC TGG GTG GCG CTC ACT 156  
 CCC ACG CTC GCG GCT AGG AAT GTC AGC GTC CCC ACT ACG 195  
 ACA ATA CGA CGC CAC GTC GAC TTG CTC GTT GGG GCG GCT 234  
 GCT TTC TGC TCC GCT ATG TAC GTG GGG GAT CTC TGC GGA 273  
 TCT GTT TTC CTC GTC TCC CAG CTG TTC ACC TTT TCG CCT 312  
 CGC CGA CAC GAG ACG GTA CAG GAC TGC AAC TGC TCA ATC 351  
 TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGG GAT 390  
 ATG ATG ATG AAC TGG TCG CCC ACA ACA GCC CTA GTG GTG 429  
 TCG CAG TTA CTC CGG ATC CCG CAA GCT ATC GTG GAC ATG 468  
 GTG GCG GGG GCC CAC TGG GGA GTC CTA GCG GGC CTT GCC 507  
 TAC TAT TCC ATG GTG GGC AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTG TTT GCC GGC GTT GAT GGG 576

20

25

## (2) INFORMATION FOR SEQ ID NO:16:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: IND5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

35

- 89 -

TAT GAA GTG CGC AAC GTG TCC GGG GTG TAC CAT GTC ACG 39  
 AAC GAC TGC TCC AAC TCA AGT ATT GTG TAT GAG GCA GCG 78  
 GAC ATG ATC ATG CAC ACT CCC GGG TGC GTG CCC TGC GTT 117  
 CGG GAG GGC AAC TCC TCT CGC TGC TGG GTA GCG CTC ACT 156  
 CCC ACT CTC GCG GCC AGG AAC GCC AGC GTC TCC ACC ACG 195  
 ACA ATA CGA CAC CAC GTC GAT TTG CTC GTT GGG GCG GCT 234  
 5 GCT TTC TGT TCC GCT ATG TAC GTG GGG GAT CTA TGC GGA 273  
 TCT GTT TTC CTC GTC TCC CAG CTG TTC ACC TTC TCA CCG 312  
 CGC CGG CAT GAG ACA GTA CAG GAC TGC AAT TGC TCC ATC 351  
 TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCC TGG GAT 390  
 ATG ATG ATG AAC TGG TCA CCT ACA GCA GCC CTA GTG GTA 429  
 TCG CAG TTG CTC CGG ATC CCA CAA GCT GTC GTG GAT ATG 468  
 GTG GCG GGG GCC CAC TGG GGA ATC CTG GCG GGC CTT GCC 507  
 10 TAC TAT TCC ATG GTA GGG AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG 576

## (2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 15  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: IND8  
 20  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TAT GAG GTG CGC AAC GTG TCC GGG GTG TAC CAT GTC ACG 39  
 AAC GAC TGC TCC AAC TCA AGT ATT GTG TAT GAG GCA GCG 78  
 GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC GTT 117  
 CGG GAG GGC AAC TTC TCT AGT TGC TGG GTA GCG CTC ACT 156  
 CCC ACT CTC GCG GCT AGG AAC GCC AGC GTC CCC ACC ACG 195  
 25 ACA ATA CGA CGC CAC GTC GAT TTG CTC GTT GGG GCG GCT 234  
 GCT TTC TGT TCC GCT ATG TAC GTG GGG GAT CTC TGC GGA 273  
 TCT GTT TTC CTT GTC TCC CAG CTG TTC ACC TTC TCA CCG 312  
 CGC CGG CAT GAG ACA GTA CAG GAC TGC AAT TGC TCC ATC 351  
 TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGG GAT 390  
 ATG ATG ATG AAC TGG TCA CCT ACA GCG GCC CTA GTG GTA 429  
 TCG CAG TTG CTC CGG ATC CCA CAA GCT GTC GTG GAT ATG 468  
 GTG GCG GGG GCC CAC TGG GGA ATC CTG GCG GGC CTT GCC 507  
 30 TAC TAT TCC ATG GTA GGG AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTC TTT GCC GGC GTT GAC GGG 576

## (2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 35

- 90 -

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: P10

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

	TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	GTG	TAC	CAT	GTC	ACG	39
	AAC	GAC	TGC	TCC	AAC	TCA	AGT	ATT	GTG	TAT	GAG	GCA	GCG	78
	GAC	ATG	ATA	ATG	CAC	ACC	CCC	GGG	TGC	GTG	CCC	TGT	GTT	117
	CGG	GAG	AAC	AAC	TCC	TCC	CGC	TGC	TGG	GTA	GCG	CTC	ACT	156
10	CCC	ACA	CTC	GCG	GCT	AGG	AAT	TCC	AGC	GTC	CCA	ACT	ACG	195
	GCA	ATA	CGA	CGC	CAT	GTC	GAT	TTG	CTC	GTT	GGG	GCG	GCT	234
	GCT	TTC	TGC	TCC	GCT	ATG	TAC	GTG	GGG	GAT	CTC	TGC	GGA	273
	TCT	GTT	CTC	CTC	GTC	TCC	CAG	CTG	TTC	ACC	TTC	TCA	CCT	312
	CGC	CGG	CAT	TGG	ACA	GTA	CAG	GAC	TGC	AAT	TGT	TCA	ATC	351
	TAT	CCT	GGC	CAC	GTA	TCA	GGT	CAC	CGC	ATG	GCT	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCG	CCC	ACA	GCA	GCC	CTA	GTG	GTG	429
	TCG	CAG	CTA	CTC	CGG	ATC	CCA	CAA	GCT	ATC	TTG	GAT	GTG	468
15	GTG	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTT	GCC	507
	TAC	TAT	TCC	ATG	GTG	GGG	AAC	TGG	GCT	AAG	GTC	TTG	ATT	546
	GTG	ATG	CTA	CTC	TTT	GCC	GGC	GTT	GAC	GGA				576

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 576 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S9

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

	TAT	GAA	GTG	CGC	AAC	GTA	TCC	GGG	GCG	TAC	CAT	GTC	ACG	39
	AAC	GAC	TGC	TCC	AAC	TCA	AGT	ATT	GTG	TAC	GAG	GCA	GCG	78
	GAC	GTG	ATC	ATG	CAT	ACC	CCC	GGG	TGT	GTA	CCC	TGC	GTT	117
	CAG	GAG	GGT	AAC	TCC	TCC	CAA	TGC	TGG	GTG	GCG	CTC	ACC	156
30	CCC	ACG	CTC	GCG	GCC	AGG	AAC	GCT	ACC	GTC	CCC	ACC	ACG	195
	ACA	ATA	CGA	CGT	CAT	GTC	GAT	TTG	CTC	GTT	GGG	GCG	GCT	234
	GTT	TTC	TGC	TCC	GCT	ATG	TAC	GTG	GGG	GAC	CTG	TGC	GGA	273
	TCT	GTT	TTC	CTC	ATC	TCC	CAG	CTG	TTC	ACC	ATC	TCG	CCC	312
	CGT	CGG	CAT	GAG	ACA	GTA	CAG	AAC	TGC	AAT	TGC	TCA	ATC	351
	TAT	CCC	GGA	CAC	GTG	ACA	GGT	CAT	CGC	ATG	GCC	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCG	CCT	ACA	ACA	GCC	CTA	GTG	GTA	429
35	TCG	CAG	CTA	CTC	CGG	ATC	CCA	CAA	GCT	GTC	ATG	GAT	ATG	468
	GTG	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTC	GCC	507



- 91 -

TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTT TTT GCT GGT GTT GAC GGG 576

## (2) INFORMATION FOR SEQ ID NO:20:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 10 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

	TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	GCG	TAC	CAT	GTC	ACG	39
	AAC	GAC	TGC	TCC	AAC	TCA	AGC	ATT	GTG	TAT	GAG	GCA	GTG	78
	GAC	GTG	ATC	CTG	CAC	ACC	CCT	GGG	TGC	GTG	CCC	TGC	GTT	117
15	CGG	GAG	AAC	AAC	TCC	TCC	CGT	TGC	TGG	GTG	GCG	CTC	ACT	156
	CCC	ACG	CTC	GCG	GCC	AGG	AAC	TCC	AGC	GTC	CCC	ACT	ACG	195
	ACA	ATA	CGA	CGT	CAC	GTC	GAT	TTG	CTC	GTT	GGG	GCG	GCT	234
	GCT	TTC	TGC	TCC	GCT	ATG	TAC	GTG	GGG	GAT	CTC	TGC	GGA	273
	TCT	GTT	TTC	CTT	GTT	TCC	CAG	CTG	TTC	ACC	TTC	TCG	CCT	312
	CGT	CGG	CAT	GAG	ACA	GTA	CAG	GAC	TGC	AAC	TGT	TCA	ATC	351
	TAT	CCC	GGC	CAC	GTA	ACA	GGT	CAC	CGC	ATG	GCT	TGG	GAT	390
20	ATG	ATG	ATG	AAC	TGG	TCG	CCT	ACA	GCA	GCC	TTA	GTG	GTA	429
	TCG	CAG	TTA	CTC	CGG	ATC	CCA	CAA	GCT	GTC	GTG	GAC	ATG	468
	GTG	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTT	GCC	507
	TAC	TAT	TCC	ATG	GTG	GGG	AAC	TGG	GCT	AAG	GTT	CTG	ATT	546
	GTG	ATG	CTA	CTC	TTT	GCC	GGC	GTT	GAC	GGG				576

## (2) INFORMATION FOR SEQ ID NO:21:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 30 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

	TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	ATG	TAC	CAT	GTC	ACG	39
	AAC	GAC	TGC	TCC	AAC	TCA	AGC	ATT	GTG	TAT	GAG	GCA	GCG	78
	GAC	ATG	ATC	ATG	CAC	ACC	CCC	GGG	TGC	GTG	CCC	TGC	GTT	117
35	CGG	GAG	AAC	AAC	TCC	TCC	CGC	TGC	TGG	GTA	GCG	CTC	ACT	156

- 92 -

CCC ACG CTC GCG GCC AGG AAC TCC AGC GTC CCC ACT ACG 195  
 ACA ATA CGA CGC CAC GTC GAT TTG CTC GTT GGG GCG GCT 234  
 GCT TTC TGC TCC GCC ATG TAC GTG GGG GAC CTC TGC GGA 273  
 TCT GTT TTC CTT GTC TCC CAG CTG TTC ACC TTC TCG CCT 312  
 CGC CGG TAT GAG ACA GTA CAG GAC TGC AAT TGC TCA ATC 351  
 TAT CCC GGC CGC GTA ACA GGT CAC CGC ATG GCT TGG GAT 390  
 5 ATG ATG ATG AAC TGG TCA CCT ACA ACA GCT CTA GTA GTA 429  
 TCG CAG TTA CTC CGG ATC CCA CAA GCT ATC GTG GAC ATG 468  
 GTG GCG GGG GCC CAC TGG GGA GTC CTA GCG GGC CTT GCC 507  
 TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT 546  
 GTT ATG CTA CTC TTT GCC GGC GTT GAC GGG 576

10 (2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TAT GAA GTG CGC AAC GTG TCC GGG GTG TAT CAT GTC ACG 39  
 AAC GAC TGT TCC AAC TCA AGC ATT GTG TAT GAG ACA GCG 78  
 20 GAC ATG ATC ATG CAT ACC CCC GGG TGC GTG CCC TGC GTT 117  
 CGG GAG GCC AAC TCC TCC CGC TGC TGG GTA GCG CTC ACT 156  
 CCC ACG CTA GCA GCC AGG AAC ACC AGC GTC CCC ACT ACG 195  
 ACA ATA CGA CGC CAC GTC GAT TTG CTC GTT GGG GCG GCT 234  
 GCT TTC TGC TCC GTT ATG TAC GTG GGG GAT CTC TGC GGA 273  
 TCT GTT TTC CTC GTC TCC CAG CTG TTC ACT TTT TCA CCT 312  
 CGC CGG CAC GAG ACA GTA CAG GAC TGC AAC TGT TCC ATC 351  
 25 TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGG GAC 390  
 ATG ATG ATG AAC TGG TCA CCT ACA GCA GCC CTG GTG GTA 429  
 TCG CAG TTA CTC CGG ATC CCA CAA GCT GTC GTG GAC ATG 468  
 GTA GCG GGG GCC CAC TGG GGA GTC CTG GCG GGC CTT GCA 507  
 TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT 546  
 GTG ATG CTA CTC TTT GCT GGC GTT GAC GGG 576

30 (2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (vi) ORIGINAL SOURCE:

- 93 -

(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: T3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

	TAC	GAA	GTG	CGC	AAC	GTG	TCC	GGG	GTG	TAC	TAT	GTC	ACG	39
5	AAC	GAC	TGT	TCC	AAC	TCA	AGC	ATT	GTG	TAT	GAG	ACA	GCG	78
	GAC	ATG	ATC	ATG	CAC	ACC	CCT	GGG	TGC	GTG	CCC	TGC	GTT	117
	CGG	GAG	AGC	AAT	TCC	TCC	CGC	TGC	TGG	GTA	GCG	CTT	ACT	156
	CCC	ACG	CTC	GCG	GCC	AGG	AAC	GCC	AGC	GTC	CCC	ACT	AAG	195
	ACA	ATA	CGA	CGT	CAC	GTG	GAC	TTG	CTC	GTT	GGG	GCG	GCT	234
	GCT	TTC	TGT	TCC	GCT	ATG	TAC	GTG	GGG	GAT	CTC	TGC	GGA	273
	TCT	GTT	TTC	CTC	GTC	TCC	CAG	CTG	TTC	ACT	TTC	TCG	CCT	312
10	CGC	CGG	CAT	GAG	ACA	GTA	CAG	GAC	TGC	AAC	TGC	TCA	ATC	351
	TAT	CCC	GGC	CAC	GTA	ACA	GGT	CAC	CGT	ATG	GCT	TGG	GAT	390
	ATG	ATG	ATG	AAC	TGG	TCG	CCC	ACA	ACG	GCA	CTA	GTG	GTG	429
	TCG	CAG	TTG	CTC	CGG	ATC	CCA	CAA	GCT	GTC	GTG	GAC	ATG	468
	GTG	GCG	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTT	GCC	507
	TAC	TAT	TCC	ATG	GTG	GGG	AAC	TGG	GCT	AAG	GTT	TTG	ATT	546
	GTG	CTG	CTA	CTC	TTT	GCC	GGC	GTT	GAT	GGG				576

15

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

25	TAT	GAA	GTG	CGC	AAC	GTG	TCC	GGG	ATG	TAC	CAT	GTC	ACG	39
	AAC	GAC	TGC	TCC	AAC	TCA	AGC	ATT	GTG	TTT	GAG	GCA	GCG	78
	GAC	TTG	ATC	ATG	CAC	ACC	CCC	GGG	TGC	GTG	CCC	TGC	GTT	117
	CGG	GAG	GGC	AAC	TCC	TCC	CGC	TGC	TGG	GTA	GCG	CTC	ACT	156
	CCC	ACG	CTC	GCG	GCC	AGG	AAC	ACC	AGC	GTC	CCC	ACT	ACG	195
	ACG	ATA	CGA	CGC	CAT	GTG	GAT	TTG	CTC	GTT	GGG	GCG	GCT	234
	GCT	TTC	TGC	TCC	GCT	ATG	TAT	GTG	GGA	GAC	CTC	TGC	GGA	273
	TCT	GTT	TTC	CTC	GTC	TCT	CAG	CTG	TTC	ACC	TTC	TCG	CCT	312
30	CGC	CGG	CAT	GAG	ACT	TTG	CAG	GAC	TGC	AAC	TGC	TCA	ATC	351
	TAT	CCC	GGC	CAT	CTG	TCA	GGT	CAC	CGC	ATG	GCT	TGG	GAC	390
	ATG	ATG	ATG	AAC	TGG	TCG	CCT	ACA	ACA	GCT	CTA	GTG	GTG	429
	TCG	CAG	TTA	CTC	CGG	ATC	CCA	CAA	GCT	GTC	ATG	GAC	ATG	468
	GTG	ACA	GGG	GCC	CAC	TGG	GGA	GTC	CTG	GCG	GGC	CTT	GCC	507
	TAC	TAT	TCC	ATG	GCG	GGG	AAC	TGG	GCT	AAG	GTT	TTA	ATT	546
	GTG	ATG	CTA	CTC	TTT	GCC	GGC	GTT	GAT	GGG				576

35

- 94 -

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US6

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

10	TAT GAA GTG CGC AAC GTG TCC GGG ATG TAC CAT GTC ACG	39
	AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAG GCA GCG	78
	GAC ATG ATC ATG CAC ACT CCC GGG TGC GTG CCC TGT GTT	117
	CGG GAG AAC AAT TCC TCC CGC TGC TGG GTA GCG CTC ACT	156
	CCC ACG CTC GCG GCC AGG AAC GCT AGC GTC CCC ACT ACG	195
	ACA ATA CGA CGC CAC GTC GAT TTG CTC GTT GGG GCG GCT	234
	ACT TTC TGC TCC GCT ATG TAC GTG GGG GAC CTC TGC GGG	273
15	TCC GTT TTC CTC ATC TCC CAG CTG TTC ACC TTC TCG CCT	312
	CGT CAG CAT GAG ACA GTA CAG GAC TGC AAT TGT TCA ATC	351
	TAT CCC GGC CAC GTA TCA GGT CAC CGC ATG GCT TGG GAT	390
	ATG ATG ATG AAT TGG TCA CCT ACA GCA GCC CTA GTG GTA	429
	TCG CAG TTA CTC CGG ATC CCA CAA GCT GTC ATG GAC ATG	468
	GTG GCG GGG GCC CAC TGG GGA GTC CTG GCG GGC CTT GCC	507
	TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT CTG ATT	546
20	GTG TTG CTA CTC TTT GCC GGC GTT GAC GGG	576

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

30	GCC CAA GTG AGG AAC ACC AGC CGC GGT TAC ATG GTG ACT	39
	AAC GAC TGT TCC AAT GAG AGC ATC ACC TGG CAG CTC CAA	78
	GCC GCG GTT CTC CAC GTC CCC GGG TGT ATC CCG TGT GAG	117
	AGG CTG GGA AAT ACA TCC CGA TGC TGG ATA CCG GTC ACA	156
	CCA AAC GTG GCC GTG CGG CAG CCC GGC GCT CTT ACG CAG	195
	GGC TTG CGG ACG CAC ATC GAC ATG GTT GTG ATG TCC GCC	234
	ACG CTC TGC TCT GCC CTC TAC GTG GGG GAC CTC TGC GGC	273
35	GGG GTG ATG CTC GCA GCC CAG ATG TTC ATT GTC TCG CCG	312

- 95 -

CGA CGC CAC TGG TTT GTG CAA GAA TGC AAT TGC TCC ATC 351  
TAC CCC GGT ACC ATC ACT GGA CAC CGT ATG GCA TGG GAC 390  
ATG ATG ATG AAC TGG TCG CCC ACA GCC ACC ATG ATC CTG 429  
GCG TAC GCG ATG CGC GTT CCC GAG GTC ATC ATA GAC ATC 468  
ATC GGC GGG GCT CAC TGG GGC GTC ATG TTT GGC TTG GCC 507  
TAC TTC TCT ATG CAG GGA GCG TGG GCG AAG GTC ATT GTC 546  
5 ATC CTC TTG CTG GCT GCT GGG GTG GAC GCG 576

## (2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 576 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: T4

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCA CAA GTG AAG AAC ACC ACT AAC AGC TAC ATG GTG ACC 39  
AAC GAC TGT TCT AAT GAC AGC ATC ACT TGG CAG CTC CAG 78  
GCC GCG GTC CTC CAC GTC CCC GGG TGT GTC CCG TGC GAG 117  
AAA ACG GGA AAT ACA TCT CGG TGC TGG ATA CCG GTT TCA 156  
CCA AAC GTG GCC GTG CGG CAG CCC GGC GCC CTC ACG CAG 195  
GGC TTG CCG ACG CAC ATT GAC ATG GTT GTG ATG TCC GCC 234  
20 ACG CTC TGC TCT GCT CTT TAC GTG GGG GAC CTC TGC GGC 273  
GGG GTG ATG CTC GCA GCC CAG ATG TTC ATC GTC TCG CCG 312  
CAA CAT CAC TGG TTT GTG CAA GAC TGC AAT TGC TCT ATC 351  
TAC CCT GGC ACC ATC ACT GGA CAC CGT ATG GCA TGG GAT 390  
ATG ATG ATG AAC TGG TCG CCC ACG GCC ACC ATG ATC CTG 429  
GCG TAC GCG ATG CGC GTT CCC GAG GTC ATC TTA GAC ATC 468  
GTT AGC GGG GCA CAC TGG GGC GTC ATG TTC GGC TTG GCC 507  
25 TAC TTC TCT ATG CAG GGA GCG TGG GCG AAA GTC GTT GTC 546  
ATC CTT CTG CTG GCC GCT GGG GTG GAC GCG 576

## (2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 576 base pairs  
(B) TYPE: nucleic acid  
30 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: T9

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

- 96 -

GCC GAA GTG AAG AAC ACC AGT ACC AGC TAC ATG GTG ACA 39  
 AAT GAC TGT TCC AAC GAC AGC ATC ACC TGG CAA CTC CAG 78  
 GCC GCG GTC CTC CAC GTC CCC GGG TGC GTC CCG TGC GAG 117  
 AGA GTT GGA AAC GCG TCG CGG TGC TGG ATA CCG GTC TCG 156  
 CCA AAC GTA GCT GTG CAG CGG CCT GGC GCC CTC ACG CAG 195  
 5 GGC TTG CGG ACG CAC ATC GAC ATG GTT GTG ATG TCC GCC 234  
 ACG CTC TGC TCC GCT CTC TAC GTG GGG GAT CTC TGC GGC 273  
 GGG GTA ATG CTC GCC GCT CAG ATG TTC ATT ATC TCG CCG 312  
 CAG CAC CAC TGG TTT GTG CAG GAA TGC AAC TGC TCC ATT 351  
 TAC CCT GGT ACC ATC ACT GGA CAC CGT ATG GCA TGG GAC 390  
 ATG ATG ATG AAC TGG TCG CCC ACA ACC ACC ATG ATC TTG 429  
 GCG TAC GCG ATG CGC GTT CCC GAG GTC ATC ATA GAC ATC 468  
 10 ATC AGC GGA GCT CAC TGG GGC GTC ATG TTC GGC CTA GCC 507  
 TAC TTC TCT ATG CAG GGA GCG TGG GCG AAG GTC GTT GTC 546  
 ATC CTG TTG CTC ACC GCT GGC GTG GAC GCG 576

## (2) INFORMATION FOR SEQ ID NO:29:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US10

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GTC CAA GTG AAA AAC ACC AGT ACC AGC TAT ATG GTG ACC 39  
 AAT GAC TGC TCC AAC GAC AGC ATC ACT TGG CAA CTT GAG 78  
 GCT GCG GTC CTC CAC GTT CCC GGG TGT GTC CCG TGC GAG 117  
 AAA GTG GGA AAT ACA TCT CGG TGC TGG ATA CCG GTC TCA 156  
 25 CCA AAT GTG GCC GTG CAG CGG CCT GGC GCC CTC ACG CAG 195  
 GGC TTG CGG ACT CAC ATC GAC ATG GTC GTG ATG TCC GCC 234  
 ACG CTC TGC TCC GCT CTT TAC GTG GGG GAC TTC TGC GGT 273  
 GGG ATG ATG CTC GCA GCC CAA ATG TTC ATT GTC TCG CCG 312  
 CGC CAC CAC TCG TTT GTG CAG GAA TGC AAC TGC TCC ATC 351  
 TAC CCC GGT ACC ATC ACC GGG CAC CGT ATG GCA TGG GAC 390  
 ATG ATG ATG AAC TGG TCG CCC ACG GCC ACT TTG ATC CTG 429  
 GCG TAC GTG ATG CGC GTT CCC GAG GTC ATC ATA GAC ATC 468  
 30 ATT AGC GGG GCG CAT TGG GGC GTC TTG TTC GGC TTA GCC 507  
 TAC TTC TCT ATG CAG GGA GCG TGG GCG AAA GTC GTT GTC 546  
 ATC CTT CTG CTA GCC GCT GGG GTG GAC GCG 576

## (2) INFORMATION FOR SEQ ID NO:30:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs

- 97 -

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

	GTG	GAA	GTC	AGG	AAC	ATC	AGT	TCC	AGC	TAC	TAC	GCC	ACC	39
	AAT	GAT	TGC	TCA	AAC	AAC	AGC	ATC	ACC	TGG	CAA	CTC	ACC	78
	GAC	GCA	GTT	CTC	CAC	CTT	CCC	GGA	TGC	GTC	CCA	TGT	GAG	117
	AAT	GAC	AAT	GGC	ACC	CTG	CGC	TGC	TGG	ATA	CAA	GTG	ACA	156
10	CCT	AAT	GTG	GCT	GTG	AAA	CAC	CGC	GGC	GCA	CTT	ACT	CAT	195
	AAC	CTG	CGA	ACA	CAC	GTC	GAC	GTG	ATC	GTA	ATG	GCA	GCT	234
	ACG	GTC	TGC	TCG	GCC	TTG	TAT	GTG	GGA	GAC	GTA	TGC	GGG	273
	GCC	GTG	ATG	ATC	GTG	TCG	CAG	GCT	CTC	ATA	ATA	TCG	CCT	312
	GAA	CGC	CAC	AAC	TTT	ACC	CAG	GAG	TGC	AAC	TGT	TCC	ATC	351
	TAC	CAA	GGT	CAT	ATC	ACC	GGC	CAC	CGC	ATG	GCA	TGG	GAC	390
	ATG	ATG	CTA	AAC	TGG	TCA	CCA	ACT	CTT	ACC	ATG	ATC	CTC	429
15	GCC	TAT	GCC	GCT	CGT	GTT	CCT	GAG	CTA	GCC	CTC	CAG	GTT	468
	GTC	TTC	GGC	GGC	CAT	TGG	GGC	GTG	GTG	TTT	GGC	TTG	GCC	507
	TAT	TTC	TCC	ATG	CAG	GGA	GCG	TGG	GCC	AAA	GTC	ATT	GCC	546
	ATC	CTC	CTT	CTT	GTC	GCA	GGA	GTG	GAT	GCA				576

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

	GTG	GAA	GTC	AGG	AAC	ACC	AGT	TCT	AGT	TAC	TAC	GCC	ACC	39
	AAT	GAT	TGC	TCA	AAC	AAC	AGC	ATC	ACC	TGG	CAA	CTC	ACC	78
	AAC	GCA	GTT	CTC	CAC	CTT	CCC	GGA	TGC	GTC	CCA	TGT	GAG	117
	AAT	GAC	AAT	GGC	ACC	CTG	CAC	TGC	TGG	ATA	CAA	GTG	ACA	156
30	CCT	AAT	GTG	GCT	GTG	AAA	CAC	CGC	GGC	GCA	CTC	ACT	CAC	195
	AAC	CTG	CGA	GCA	CAT	ATA	GAT	ATG	ATT	GTA	ATG	GCA	GCT	234
	ACG	GTC	TGC	TCG	GCC	TTG	TAT	GTG	GGA	GAC	GTG	TGC	GGG	273
	GCC	GTG	ATG	ATC	GTG	TCG	CAG	GCT	TTC	ATA	GTA	TCG	CCA	312
	GAA	CAC	CAC	CAC	TTT	ACC	CAA	GAG	TGC	AAC	TGT	TCC	ATC	351
	TAC	CAA	GGT	CAC	ATC	ACC	GGC	CAC	CGC	ATG	GCA	TGG	GAC	390
	ATG	ATG	CTT	AAC	TGG	TCA	CCA	ACT	CTC	ACC	ATG	ATC	CTC	429
35	GCC	TAT	GCC	CGC	GTT	CCT	GAG	CTA	GTC	CTT	GAA	GTC		468
	GTC	TTC	GGT	GGT	CAT	TGG	GGT	GTG	GTG	TTT	GGC	TTG	GCC	507

- 98 -

TAT TTC TCC ATG CAG GGA GCG TGG GCC AAG GTC ATT GCC 546  
 ATC CTC CTT CTT GTA GCA GGA GTG GAT GCA 576

## (2) INFORMATION FOR SEQ ID NO:32:

5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 10 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW3

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GTG GAA GTC AGG AAC ATC AGT TCT AGC TAC TAT GCC ACC 39  
 AAT GAT TGC TCA AAC AGC AGC ATC ACC TGG CAA CTC ACC 78  
 AAC GCA GTC CTC CAC CTT CCC GGA TGC GTC CCG TGT GAG 117  
 15 AAT GAT AAT GGC ACC CTG CAC TGC TGG ATA CAA GTG ACA 156  
 CCT AAT GTG GCT GTG AAA CAC CGC GGC GCG CTC ACT CAC 195  
 AAC CTG CGA GCA CAC GTC GAT ATG ATC GTA ATG GCA GCT 234  
 ACG GTC TGC TCG GCC TTG TAT GTG GGA GAC ATG TGC GGG 273  
 GCC GTG ATG ATC GTG TCG CAG GCT TTC ATA ATA TCG CCA 312  
 GAA CGC CAC AAC TTT ACC CAA GAG TGC AAC TGT TCC ATC 351  
 TAC CAA GGT CGT ATC ACC GGC CAC CGC ATG GCG TGG GAC 390  
 ATG ATG CTA AAC TGG TCA CCA ACT CTT ACC ATG ATC CTT 429  
 20 GCC TAT GCC GCT CGT GTT CCT GAG CTA GTC CTT GAA GTT 468  
 GTC TTC GGC GGC CAT TGG GGC GTG GTG TTT GGC TTG GCC 507  
 TAT TTC TCC ATG CAA GGA GCG TGG GCC AAG GTC ATT GCC 546  
 ATC CTC CTG CTT GTC GCA GGA GTG GAT GCA 576

## (2) INFORMATION FOR SEQ ID NO:33:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 30 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GTG GAA GTT AGA AAC ACC AGT TTT AGC TAC TAC GCC ACC 39  
 AAT GAT TGC TCG AAC AAC AGC ATC ACC TGG CAG CTC ACC 78  
 AAC GCA GTT CTC CAC CTT CCC GGA TGC CCA TGT GAG 117  
 35 AAT GAC AAT GGC ACC TTG CGC TGC TGG ATA CAA GTA ACA 156



- 99 -

CCT AAT GTG GCT GTG AAA CAC CGT GGC GCA CTC ACT CAC 195  
 AAC CTG CGA ACG CAT GTC GAC GTG ATC GTA ATG GCA GCT 234  
 ACG GTC TGC TCG GCC TTG TAT GTG GGG GAC GTG TGC GGG 273  
 GCC GTG ATG ATA GCG TCG CAG GCT TTC ATA ATA TCG CCA 312  
 GAA CGC CAC AAC TTC ACC CAG GAG TGC AAC TGT TCC ATC 351  
 TAC CAA GGT CAT ATC ACC GGC CAC CGC ATG GCA TGG GAC 390  
 5 ATG ATG CTG AAC TGG TCA CCA ACT CTC ACC ATG ATC CTC 429  
 GCC TAC GCT GCT CGT GTG CCT GAA CTA GTC CTT GAA GTT 468  
 GTC TTC GGC GGC CAT TGG GGC GTG GTG TTT GGC TTG GCC 507  
 TAT TTC TCC ATG CAA GGA GCG TGG GCC AAA GTC ATC GCC 546  
 ATC CTC CTC CTT GTC GCA GGA GTG GAC GCA 576

10 (2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S83

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GTG GAG GTC AAG GAC ACC GGC GAC TCC TAC ATG CCG ACC 39  
 AAC GAT TGC TCC AAC TCT AGT ATC GTT TGG CAG CTT GAA 78  
 20 GGA GCA GTG CTT CAT ACT CCT GGA TGC GTC CCT TGT GAG 117  
 CGT ACC GCC AAC GTC TCT CGA TGT TGG GTG CCG GTT GCC 156  
 CCC AAT CTC GCC ATA AGT CAA CCT GGC GCT CTC ACT AAG 195  
 GGC CTG CGA GCA CAC ATC GAT ATC ATC GTG ATG TCT GCT 234  
 ACG GTC TGT TCT GCC CTT TAT GTG GGG GAC GTG TGT GGC 273  
 GCG CTG ATG CTG GCC GCT CAG GTC GTC GTC GTG TCG CCA 312  
 CAA CAC CAT ACG TTT GTC CAG GAA TGC AAC TGT TCC ATA 351  
 25 TAC CCG GGC CGC ATT ACG GGA CAC CGC ATG GCT TGG GAT 390  
 ATG ATG ATG AAC TGG TCG CCC ACT ACC ACC ATG CTC CTG 429  
 GCG TAC TTG GTG CGC ATC CCG GAA GTC ATC TTG GAT ATT 468  
 GTT ACA GGA GGT CAT TGG GGT GTA ATG TTT GGC CTC GCT 507  
 TAC TTC TCC ATG CAG GGA TCG TGG GCG AAG GTC ATC GTT 546  
 ATC CTC CTG CTG ACT GCT GGG GTG GAG GCG 576

30 (2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (vi) ORIGINAL SOURCE:

- 100 -

(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: DK12

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

5	TTA	GAG	TGG	CGG	AAT	GTG	TCC	GGC	CTC	TAC	GTC	CTT	ACC	39
	AAC	GAC	TGT	TCC	AAT	AGC	AGT	ATC	GTG	TAT	GAG	GCC	GAT	78
	GAC	GTC	ATT	CTG	CAC	ACA	CCT	GGC	TGT	GTA	CCT	TGT	GTT	117
	CAG	GAC	GGC	AAT	ACA	TCT	ACG	TGC	TGG	ACC	TCA	GTG	ACG	156
	CCT	ACA	GTG	GCA	GTC	AGG	TAC	GTC	GGA	GCA	ACC	ACC	GCT	195
	TCG	ATA	CGC	AGT	CAT	GTG	GAC	CTG	CTA	GTG	GGC	GCG	GCC	234
	ACG	ATG	TGC	TCT	GCG	CTC	TAC	GTG	GGT	GAT	GTG	TGT	GGG	273
	GCC	GTC	TTC	CTT	GTG	GGA	CAA	GCC	TTC	ACG	TTC	AGA	CCT	312
10	CGT	CGC	CAT	CAA	ACA	GTC	CAG	ACC	TGT	AAC	TGC	TCG	CTG	351
	TAC	CCA	GGC	CAT	CTT	TCA	GGA	CAT	CGA	ATG	GCT	TGG	GAT	390
	ATG	ATG	ATG	AAT	TGG	TCC	CCC	GCT	GTG	GGT	ATG	GTG	GTA	429
	GCG	CAC	GTC	CTG	CGT	CTG	CCC	CAG	ACC	TTG	TTC	GAC	ATA	468
	ATA	GCT	GGG	GCC	CAT	TGG	GGC	ATC	ATG	GCG	GGC	CTA	GCC	507
	TAT	TAC	TCC	ATG	CAG	GGC	AAC	TGG	GCC	AAG	GTC	GCT	ATC	546
	ATC	ATG	GTT	ATG	TTT	TCA	GGA	GTC	GAT	GCC				576

15

## (2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 576 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: HK10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

25	CTA	GAG	TGG	CGG	AAT	GTG	TCT	GGC	CTC	TAT	GTC	CTT	ACC	39
	AAC	GAC	TGT	CCC	AAT	AGC	AGT	ATT	GTG	TAT	GAG	GCC	GAT	78
	GAC	GTC	ATT	CTG	CAC	ACA	CCT	GGC	TGT	GTA	CCT	TGT	GTT	117
	CAG	GAC	GGC	AAT	ACA	TCC	ACG	TGC	TGG	ACC	TCG	GTG	ACA	156
	CCT	ACA	GTG	GCA	GTC	AGG	TAC	GTC	GGA	GCA	ACC	ACC	GCC	195
	TCG	ATA	CGC	AGT	CAT	GTG	GAC	CTG	TTA	GTG	GGC	GCG	GCC	234
	ACG	ATG	TGC	TCT	GCG	CTC	TAC	GTG	GGC	GAT	ATG	TGT	GGG	273
	GCC	GTC	TTC	CTC	GTG	GGA	CAA	GCC	TTC	ACG	TTC	AGA	CCG	312
30	CGT	CGC	CAT	CAA	ACG	GTC	CAG	ACC	TGT	AAC	TGC	TCG	CTG	351
	TAC	CCA	GGC	CAC	CTT	TCA	GGA	CAT	CGA	ATG	GCT	TGG	GAT	390
	ATG	ATG	ATG	AAT	TGG	TCC	CCC	GCC	GTG	GGT	ATG	GTG	GTG	429
	GCG	CAC	GTC	CTG	CGG	TTG	CCC	CAG	ACC	TTG	TTC	GAC	ATA	468
	ATA	GCC	GGG	GCC	CAT	TGG	GGC	ATC	TTG	GCA	GGC	CTA	GCC	507
	TAT	TAC	TCC	ATG	CAG	GGC	AAC	TGG	GCC	AAG	GTC	GCT	ATC	546
	ATC	ATG	GTT	ATG	TTT	TCA	GGG	GTC	GAT	GCC				576

35

- 101 -

## (2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

10	CTA GAG TGG CGG AAT ACG TCT GGC CTC TAT GTC CTC ACC	39
	AAC GAC TGT TCC AAT AGC AGT ATT GTG TAT GAG GCC GAT	78
	GAC GTT ATT CTG CAC ACA CCT GGC TGT GTA CCT TGT GTT	117
	CAG GAC GGT AAT ACA TCC ACG TGC TGG ACC CCA GTG ACA	156
	CCT ACA GTG GCA GTC AGG TAT GTC GGA GCA ACC ACC GCT	195
	TCG ATA CGC AGT CAT GTG GAC CTA TTG GTG GGC GCG GCC	234
	ACT ATG TGC TCT GCG CTC TAC GTG GGT GAT ATG TGT GGG	273
15	GCC GTC TTT CTC GTG GGA CAA GCC TTC ACG TTC AGA CCT	312
	CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC TCG CTG	351
	TAC CCA GGC CAT CTT TCA GGA CAT CGC ATG GCT TGG GAT	390
	ATG ATG ATG AAT TGG TCC CCC GCT GTG GGT ATG GTG GTG	429
	GCG CAC GTT CTG CGT TTG CCC CAG ACC GTG TTC GAC ATA	468
	ATA GCC GGG GCC CAT TGG GGC ATC TTG GCG GGC CTA GCC	507
	TAT TAC TCC ATG CAA GGC AAC TGG GCC AAG GTC GCT ATC	546
20	ATC ATG GTT ATG TTT TCA GGG GTC GAC GCC	576

## (2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S52

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

30	CTA GAG TGG CGG AAT ACG TCT GGC CTC TAT GTC CTT ACC	39
	AAC GAC TGT TCC AAT AGC AGT ATT GTG TAT GAG GCC GAT	78
	GAC GTC ATT CTG CAC ACA CCC GGC TGT GTA CCT TGT GTT	117
	CAG GAC GGC AAT ACA TCC ATG TGC TGG ACC CCA GTG ACA	156
	CCT ACG GTG GCA GTC AGG TAC GTC GGA GCA ACC ACC GCT	195
	TCG ATA CGC AGT CAT GTG GAC CTA TTA GTG GGC GCG GCC	234
	ACG CTG TGC TCT GCG CTC TAT GTG GGT GAT ATG TGT GGG	273
35	GCC GTC TTT CTC GTG GGA CAA GCC TTC ACG TTC AGA CCT	312

- 102 -

CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC TCG CTG 351  
TAC CCA GGC CAT GTT TCA GGA CAT CGA ATG GCT TGG GAT 390  
ATG ATG ATG AAT TGG TCC CCC GCT GTG GGT ATG GTG GTG 421  
GCG CAC ATC CTG CGA TTG CCC CAG ACC TTG TTT GAC ATA 468  
CTG GCC GGG GCC CAT TGG GGC ATC TTG GCG GGC CTA GCC 507  
TAT TAT TCT ATG CAG GGC AAC TGG GCC AAG GTC GCT ATT 546  
5 GTC ATG ATT ATG TTT TCA GGG GTC GAT GCC 576

## (2) INFORMATION FOR SEQ ID NO:39:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 576 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: S54

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CTA GAG TGG CGG AAT ACG TCT GGC CTC TAT ATC CTT ACC 39  
AAC GAC TGT TCC AAT AGC AGT ATT GTG TAT GAG GCC GAT 78  
GAC GTC ATT CTG CAC ACA CCC GGC TGT GTA CCT TGT GTT 117  
CAG GAC GGC AAT ACA TCC ACG TGC TGG ACC CCA GTG ACA 156  
CCT ACG GTG GCA GTC AGG TAC GTC GGA GCA ACC ACC GCT 195  
20 TCG ATA CGC AGT CAT GTG GAC CTA TTA GTG GGC GCG GCC 234  
ACG CTG TGC TCT GCG CTC TAT GTG GGT GAT ATG TGT GGG 273  
GCC GTC TTT CTC GTG GGA CAA GCC TTC ACG TTC AGA CCT 312  
CGT CGC CAT CAA ACG GTC CAG ACC TGT AAC TGC TCG CTG 351  
TAC CCA GGC CAT CTT TCA GGA CAT CGA ATG GCT TGG GAT 390  
ATG ATG ATG AAT TGG TCC CCC GCT GTG GGT ATG GTG GTG 429  
GCG CAC ATC CTG CGA TTG CCC CAG ACC TTG TTT GAC ATA 468  
CTG GCC GGG GCC CAT TGG GGC ATC TTG GCG GGC CTA GCC 507  
25 TAT TAT TCT ATG CAG GGC AAC TGG GCC AAG GTC GCT ATC 546  
ATC ATG ATT ATG TTT TCA GGG GTC GAT GCC 576

## (2) INFORMATION FOR SEQ ID NO:40:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 576 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: Z4

## 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

- 103 -

	GAG	CAC	TAC	CGG	AAT	GCT	TCG	GGC	ATC	TAT	CAC	ATC	ACC	39
	AAT	GAT	TGT	CCG	AAT	TCC	AGT	ATA	GTC	TAT	GAA	GCT	GAC	78
	CAT	CAC	ATC	CTA	CAC	TTG	CCG	GGG	TGC	GTA	CCC	TGT	GTG	117
	ATG	ACT	GGG	AAC	ACA	TCG	CGT	TGC	TGG	ACG	CCG	GTG	ACG	156
	CCT	ACA	GTG	GCT	GTC	GCA	CAC	CCG	GGC	GCT	CCG	CTT	GAG	195
5	TCG	TTC	CGG	CGA	CAT	GTG	GAC	TTA	ATG	GTA	GGC	GCG	GCC	234
	ACT	TTG	TGT	TCT	GCC	CTC	TAT	GTT	GGG	GAC	CTC	TGC	GGA	273
	GGT	GCC	TTC	CTG	ATG	GGG	CAG	ATG	ATC	ACT	TTT	CGG	CCG	312
	CGT	CGC	CAC	TGG	ACC	ACG	CAG	GAG	TGC	AAT	TGT	TCC	ATC	351
	TAC	ACT	GGC	CAT	ATC	ACC	GGC	CAC	AGG	ATG	GCG	TGG	GAC	390
	ATG	ATG	ATG	AAC	TGG	AGC	CCT	ACC	ACC	ACT	CTG	CTC	CTC	429
	GCC	CAG	ATC	ATG	AGG	GTC	CCC	ACA	GCC	TTT	CTC	GAC	ATG	468
10	GTT	GCC	GGA	GGC	CAC	TGG	GGC	GTC	CTC	GCG	GGC	TTG	GCG	507
	TAC	TTC	AGC	ATG	CAA	GGC	AAT	TGG	GCC	AAG	GTA	GTC	CTG	546
	GTC	CTT	TTC	CTC	TTT	GCT	GGG	GTA	GAC	GCC				576

## (2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 576 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: Z1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

	GTG	CAC	TAC	CGG	AAT	GCT	TCG	GGC	GTC	TAT	CAT	GTC	ACC	39
	AAT	GAT	TGC	CCT	AAC	ACC	AGC	ATA	GTG	TAC	GAG	ACG	GAG	78
	CAC	CAC	ATC	ATG	CAC	TTG	CCA	GGG	TGT	GTC	CCC	TGT	GTG	117
	CGG	ACG	GAG	AAT	ACT	TCT	CGC	TGC	TGG	GTG	CCC	TTG	ACC	156
25	CCC	ACT	GTG	GCC	GCG	CCC	TAT	CCC	AAC	GCA	CCG	TTA	GAG	195
	TCC	ATG	CGC	AGG	CAT	GTA	GAC	CTG	ATG	GTG	GGT	GCG	GCT	234
	ACT	ATG	TGT	TCC	GCC	TTC	TAC	ATT	GGA	GAT	CTG	TGT	GGA	273
	GGC	GTC	TTC	CTA	GTG	GGC	CAG	CTG	TTC	GAC	TTC	CGA	CCG	312
	CGC	CGG	CAC	TGG	ACC	ACC	CAG	GAT	TGC	AAC	TGC	TCC	ATC	351
	TAT	CCT	GGT	CAC	GTC	TCG	GGC	CAC	AGG	ATG	GCC	TGG	GAC	390
	ATG	ATG	ATG	AAC	TGG	AGC	CCT	ACC	AGC	GCG	CTG	ATT	ATG	429
	GCT	CAG	ATC	TTA	CGG	ATC	CCC	TCT	ATC	CTA	GGT	GAC	TTG	468
30	CTC	ACC	GGG	GGT	CAC	TGG	GGA	GTT	CTT	GCT	GGT	CTA	GCT	507
	TTC	TTC	AGC	ATG	CAG	AGT	AAC	TGG	GCG	AAG	GTC	ATC	CTG	546
	GTC	CTA	TTC	CTC	TTT	GCC	GGG	GTC	GAG	GGA				576

## (2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 576 base pairs

- 104 -

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

	GTT	AAC	TAT	CGC	AAT	GCC	TCG	GGC	GTC	TAT	CAC	GTC	ACC	39
	AAC	GAC	TGC	CCG	AAC	TCG	AGC	ATA	GTG	TAT	GAG	GCC	GAA	78
	CAC	CAG	ATC	TTA	CAC	CTC	CCA	GGG	TGC	TTG	CCC	TGT	GTG	117
10	AGG	GTT	GGG	AAT	CAG	TCA	CGC	TGC	TGG	GTG	GCC	CTT	ACT	156
	CCC	ACC	GTG	GCG	GTG	TCT	TAT	ATC	GGT	GCT	CCG	CTT	GAC	195
	TCC	CTC	CGG	AGA	CAT	GTG	GAC	CTG	ATG	GTG	GGC	GCC	GCT	234
	ACT	GTA	TGC	TCT	GCC	CTC	TAC	GTT	GGA	GAT	CTG	TGC	GGT	273
	GGT	GCA	TTC	TTG	GTT	GGC	CAG	ATG	TTC	TCC	TTC	CAG	CCG	312
	CGA	CGC	CAC	TGG	ACT	ACG	CAG	GAC	TGC	AAT	TGT	TCT	ATC	351
	TAC	GCA	GGG	CAT	ATC	ACG	GGC	CAC	AGG	ATG	GCA	TGG	GAC	390
	ATG	ATG	ATG	AAC	TGG	AGT	CCC	ACA	ACC	ACC	CTG	CTT	CTC	429
15	GCC	CAG	GTC	ATG	AGG	ATC	CCT	AGC	ACT	CTG	GTA	GAT	CTA	468
	CTC	GCT	GGA	GGG	CAC	TGG	GGC	GTC	CTT	GTT	GGG	TTG	GCG	507
	TAC	TTC	AGT	ATG	CAA	GCT	AAT	TGG	GCC	AAA	GTC	ATC	CTG	546
	GTC	CTT	TTC	CTC	TTC	GCT	GGA	GTT	GAT	GCC				576

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

	GTC	AAC	TAT	CAC	AAT	GCC	TCG	GGC	GTC	TAT	CAC	ATC	ACC	39
	AAC	GAC	TGC	CCG	AAC	TCG	AGC	ATA	ATG	TAT	GAG	GCC	GAA	78
	CAC	CAC	ATC	CTA	CAC	CTC	CCA	GGG	TGC	GTA	CCC	TGT	GTG	117
30	AGG	GAG	GGG	AAC	CAG	TCA	CGC	TGC	TGG	GTG	GCC	CTT	ACT	156
	CCC	ACC	GTG	GCG	GCG	CCT	TAT	ATC	GGT	GCA	CCG	CTT	GAA	195
	TCC	ATC	CGG	AGA	CAT	GTG	GAC	CTG	ATG	GTA	GGC	GCT	GCT	234
	ACA	GTG	TGC	TCC	GCT	CTC	TAC	ATT	GGG	GAC	CTG	TGC	GGT	273
	GGC	GTA	TTT	TTG	GTT	GGT	CAG	ATG	TTT	TCT	TTC	CAG	CCG	312
	CGA	CGC	CAC	TGG	ACT	ACG	CAG	GAC	TGC	AAT	TGT	TCC	ATC	351
	TAT	GCG	GGG	CAC	GTT	ACA	GGC	CAC	AGA	ATG	GCA	TGG	GAC	390
35	ATG	ATG	ATG	AAC	TGG	AGT	CCC	ACA	ACC	ACC	TTG	GTC	CTC	429
	GCC	CAG	GTT	ATG	AGG	ATC	CCT	AGC	ACT	CTG	GTG	GAC	CTA	468

- 105 -

CTC	ACT	GGA	GGG	CAC	TGG	GGT	ATC	CTT	ATC	GGG	GTG	GCA	507
TAC	TTC	TGC	ATG	CAA	GCT	AAT	TGG	GCC	AAG	GTC	ATT	CTG	546
GTC	CTT	TTC	CTC	TAC	GCT	GGA	GTT	GAT	GCC				576

## (2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK13

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TAC	AAC	TAT	CGC	AAC	AGC	TCG	GGT	GTC	TAC	CAT	GTC	ACC	39
AAC	GAT	TGC	CCG	AAC	TCG	AGC	ATA	GTC	TAT	GAA	ACC	GAT	78
TAC	CAC	ATC	TTA	CAC	CTC	CCG	GGA	TGC	GTT	CCT	TGC	GTG	117
AGG	GAA	GGG	AAC	AAG	TCT	ACA	TGC	TGG	GTG	TCT	CTC	ACC	156
CCC	ACC	GTG	GCT	GCG	CAA	CAT	CTG	AAT	GCT	CCG	CTT	GAG	195
TCT	TTG	AGA	CGT	CAC	GTG	GAT	CTG	ATG	GTG	GGC	GGC	GCC	234
ACT	CTC	TGC	TCC	GCC	CTC	TAC	ATC	GGA	GAC	GTG	TGT	GGG	273
GGT	GTG	TTT	TTG	GTC	GGT	CAA	CTG	TTC	ACC	TTC	CAA	CCT	312
CGC	CGC	CAC	TGG	ACC	ACC	CAA	GAC	TGC	AAT	TGT	TCC	ATC	351
TAC	ACA	GGA	CAT	ATC	ACA	GGA	CAC	AGA	ATG	GCT	TGG	GAC	390
ATG	ATG	ATG	AAT	TGG	AGC	CCC	ACT	GCG	ACG	CTG	GTC	CTC	429
GCC	CAA	CTT	ATG	AGG	ATC	CCA	GGC	GCC	ATG	GTC	GAC	CTG	468
CTT	GCA	GGC	GGC	CAC	TGG	GGC	ATT	CTG	GTT	GGC	ATA	GCG	507
TAC	TTC	AGC	ATG	CAA	GCT	AAT	TGG	GCC	AAG	GTT	ATC	CTG	546
GTC	CTG	TTT	CTC	TTT	GCT	GGA	GTC	GAC	GCT				576

## (2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GTT	CCC	TAC	CGG	AAT	GCC	TCT	GGG	GTT	TAC	CAT	GTC	ACC	39
AAT	GAC	TGC	CCA	AAC	TCC	TCC	ATA	GTC	TAC	GAG	GCT	GAT	78
AGC	CTG	ATC	TTG	CAC	GCA	CCT	GGC	TGC	GTG	CCC	TGT	GTC	117

- 106 -

AGG CAA GAT AAT GTC AGT AGG TGC TGG GTC CAA ATC ACC 156  
 CCC ACA CTG TCA GCC CCG ACC TTC GGA GCG GTC ACG GCT 195  
 CCT CTT CGG AGG GCC GTT GAC TAC TTA GCG GGA GGA GCT 234  
 GCT CTC TGC TCC GCA CTA TAC GTC GGC GAC GCG TGC GGG 273  
 GCA GTG TTT CTG GTA GGC CAA ATG TTC ACC TAT AGG CCT 312  
 CGC CAG CAT ACC ACA GTG CAG GAC TGC AAC TGT TCC ATT 351  
 5 TAC AGT GGC CAT ATC ACC GGC CAC CGG ATG GCT TGG GAC 390  
 ATG ATG ATG AAT TGG TCA CCT ACG ACA GCC TTG CTG ATG 429  
 GCC CAG ATG CTA CGG ATC CCC CAG GTG GTC ATA GAC ATC 468  
 ATA GCC GGG GGC CAC TGG GGG GTC TTG TTT GCC GCC GCA 507  
 TAC TTT GCG TCG GCC GCC AAC TGG GCT AAG GTA GTG CTG 546  
 GTT CTG TTC CTG TTT GCG GGG GTC GAT GGC 576

10 (2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

GTT CCC TAC CGA AAC GCC TCT GGG GTT TAT CAT GTC ACC 39  
 20 AAT GAT TGC CCA AAC TCT TCC ATA GTT TAC GAG GCT GAT 78  
 AAC CTG ATC TTG CAT GCA CCT GGT TGC GTG CCT TGT GTC 117  
 AGG CAA GAT AAT GTC AGT AAG TGC TGG GTC CAA ATC ACC 156  
 CCC ACG TTG TCA GCC CCG AAT CTC GGA GCG GTC ACG GCT 195  
 CCT CTT CGG AGG GCC GTT GAC TAC TTA GCG GGA GGG GCT 234  
 GCC CTC TGC TCC GCA CTA TAC GTC GGG GAC GCG TGC GGG 273  
 GCA GTG TTT TTG GTA GGC CAA ATG TTC ACC TAT AGG CCT 312  
 25 CGC CAG CAC ACT ACG GTG CAA GAC TGC AAT TGC TCT ATT 351  
 TAC AGT GGC CAT ATC ACC GGC CAC CGG ATG GCA TGG GAC 390  
 ATG ATG ATG AAT TGG TCA CCT ACG ACG GCC TTG CTG ATG 429  
 GCC CAG TTG CTA CGG ATT CCC CAG GTG GTC ATC GAC ATC 468  
 ATT GCC GGG GGC CAC TGG GGG GTC TTG TTT GCC GCC GCA 507  
 TAT TTC GCG TCA GCG GCT AAC TGG GCT AAG GTT ATA CTG 546  
 GTC TTG TTT CTG TTT GCG GGG GTC GAT GCC 576

30 (2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (vi) ORIGINAL SOURCE:



- 107 -

(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: SAS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

	GTC	CCC	TAC	CGA	AAT	GCC	TCT	GGG	GTT	TAT	CAT	GTC	ACC	39
5	AAT	GAT	TGC	CCA	AAC	TCT	TCC	ATA	GTC	TAC	GAG	GCT	GAT	78
	AAC	CTG	ATT	CTG	CAC	GCA	CCT	GGT	TGC	GTG	CCC	TGT	GTC	117
	AAG	GAA	GGT	AAT	GTC	AGT	AGG	TGC	TGG	GTC	CAA	ATC	ACC	156
	CCC	ACA	TTG	TCA	GCC	CCG	AAC	CTC	GGA	GCG	GTC	ACG	GCT	195
	CCT	CTT	CGG	AGG	GTC	GTT	GAC	TAC	TTA	GCG	GGA	GGG	GCT	234
	GCC	CTC	TGC	TCC	GCA	CTA	TAC	GTC	GGG	GAC	GCG	TGC	GGG	273
	GCA	GTG	TTC	TTG	GTA	GGC	CAA	ATG	TTC	ACC	TAT	AGG	CCT	312
10	CGC	CAG	CAT	ACT	ACG	GTG	CAG	GAC	TGC	AAC	TGT	TCC	ATT	351
	TAC	AGC	GGC	CAT	ATC	ACC	GGC	CAC	CGA	ATG	GCA	TGG	GAC	390
	ATG	ATG	ATG	AAT	TGG	TCA	CCT	ACG	ACA	GCC	TTG	GTG	ATG	429
	GCC	CAG	GTG	CTA	CGG	ATT	CCC	CAA	GTG	GTC	ATT	GAC	ATC	468
	ATT	GCC	GGG	GGC	CAC	TGG	GGG	GTC	TTG	TTC	GCC	GTC	GCA	507
	TAC	TTC	GCG	TCA	GCG	GCT	AAC	TGG	GCT	AAG	GTT	GTG	CTG	546
	GTC	CTG	TTT	CTG	TTT	GCG	GGG	GTC	GAT	GGC				576

15

## (2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 576 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: SA6

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

25	GTT	CCT	TAC	CGG	AAT	GCC	TCT	GGG	GTG	TAT	CAT	GTT	ACC	39
	AAT	GAT	TGC	CCA	AAC	TCT	TCC	ATA	GTC	TAT	GAG	GCT	GAT	78
	GAC	CTG	ATC	CTA	CAC	GCA	CCT	GGC	TGC	GTG	CCC	TGT	GTC	117
	CGG	AAG	GAT	AAT	GTC	AGT	AGA	TGC	TGG	GTT	CAT	ATC	ACC	156
	CCC	ACA	CTA	TCA	GCC	CCG	AGC	CTC	GGA	GCG	GTC	ACG	GCT	195
	CCT	CTT	CGG	AGG	GCC	GTT	GAT	TAC	TTG	GCG	GGA	GGG	GCC	234
	GCC	CTG	TGC	TCC	GCG	TTA	TAC	GTC	GGA	GAC	GTG	TGC	GGG	273
	GCA	TTG	TTT	TTG	GTA	GGC	CAA	ATG	TTC	ACC	TAT	AGG	CCT	312
30	CGC	CAG	CAT	GCT	ACG	GTA	CAG	GAC	TGC	AAC	TGC	TCC	ATT	351
	TAC	AGT	GGC	CAT	ATC	ACT	GGC	CAC	CGG	ATG	GCA	TGG	GAC	390
	ATG	ATG	ATG	AAT	TGG	TCA	CCC	GCG	ACA	GCC	TTG	GTG	ATG	429
	GCC	CAA	ATG	CTA	CGG	ATT	CCC	CAG	GTG	GTC	ATT	GAC	ATC	468
	ATT	GCC	GGG	GGC	CAC	TGG	GGG	GTC	TTG	TTC	GCC	GCT	GCA	507
	TAC	TTC	GCG	TCG	GCG	GCT	AAC	TGG	GCT	AAG	GTT	GTG	CTG	546
	GTC	TTG	TTT	CTG	TTT	GCG	GGG	GTT	GAT	GCC				576

35

- 108 -

## (2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA7

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

10	GTC CCC TAC CGA AAT GCC TCC GGG GTT TAT CAT GTC ACC	39
	AAT GAT TGC CCG AAC TCT TCC ATA GTC TAT GAG GCT GAC	78
	AAC CTG ATC CTG CAC GCA CCT GGT TGC GTG CCC TGT GTC	117
	AGA CAA AAT AAT GTC AGT AGG TGC TGG GTC CAA ATC ACC	156
	CCC ACA TTG TCA GCC CCG AAC CTC GGA GCG GTC ACG GCT	195
	CCT CTT CGG AGG GCC GTT GAC TAC CTA GCG GGA GGG GCT	234
	GCC CTC TGC TCC GCG CTA TAC GTC GGG GAC GCG TGC GGG	273
15	GCA GTG TTT TTG GTA GGC CAG ATG TTC AGC TAT AGG CCT	312
	CGC CAG CAC ACT ACG GTG CAG GAC TGC AAC TGT TCC ATT	351
	TAC AGT GGC CAT ATC ACC GGC CAC CGA ATG GCA TGG GAC	390
	ATG ATG ATG AAT TGG TCA CCT ACG ACA GCC TTG GTG ATG	429
	GCC CAG TTG CTA CGG ATT CCC CAG GTG GTC ATC GAC ATC	468
	ATT GCC GGG GGC CAC TGG GGG GTC TTG TTC GCC GCC GCA	507
	TAT TTC GCG TCA GCG GCT AAC TGG GCT AAG GTT GTG CTG	546
20	GTC TTG TTT CTG TTT GCG GGG GTC GAT GCC	576

## (2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA13

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

30	GTT CCC TAC CGA AAT GCC TCT GGG GTT TAT CAT GTC ACC	39
	AAT GAT TGC CCA AAC TCT TCC ATC GTC TAC GAG GCT GAT	78
	GAC CTG ATC TTA CAC GCA CCT GGT TGC GTG CCC TGT GTT	117
	AGG CAG GGT AAT GTC AGT AGG TGC TGG GTC CAG ATC ACC	156
	CCC ACA CTG TCA GCC CCG AGC CTC GGA GCG GTC ACG GCT	195
	CCT CTT CGG AGG GCC GTT GAC TAC TTA GCG GGG GGG GCT	234
	GCC CTT TGC TCC GCG TTA TAC GTC GGA GAC GCG TGC GGG	273
35	GCA GTG TTT TTG GTA GGT CAA ATG TTC ACC TAT AGC CCT	312
	CGC CGG CAT AAT GTT GTG CAG GAC TGC AAC TGT TCC ATT	351

- 109 -

.

5

TAC	AGT	GGC	CAC	ATC	ACC	GGC	CAC	CGG	ATG	GCA	TGG	GAC	390
ATG	ATG	ATG	AAT	TGG	TCA	CCT	ACA	ACA	GCT	TTG	GTG	ATG	429
GCC	CAG	TTG	TTA	CGG	ATT	CCC	CAG	GTG	GTC	ATT	GAC	ATC	468
ATT	GCC	GGG	GCC	CAC	TGG	GGG	GTC	TTG	TTC	GCC	GCC	GCA	507
TAC	TAC	GCG	TCG	GCG	GCT	AAC	TGG	GCC	AAG	GTT	GTG	CTG	546
GTC	CTG	TTT	CTG	TTT	GCG	GGG	GTC	GAT	GCC				576

## (2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 576 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

15

CTT	ACC	TAC	GGC	AAC	TCC	AGT	GGG	CTA	TAC	CAT	CTC	ACA	39
AAT	GAT	TGC	CCC	AAC	TCC	AGC	ATC	GTG	CTG	GAG	GCG	GAT	78
GCT	ATG	ATC	TTG	CAT	TTG	CCT	GGA	TGC	TTG	CCT	TGT	GTG	117
AGG	GTC	GAT	GAT	CGG	TCC	ACC	TGT	TGG	CAT	GCT	GTG	ACC	156
CCC	ACC	CTG	GCC	ATA	CCA	AAT	GCT	TCC	ACG	CCC	GCA	ACG	195
GGA	TTC	CGC	AGG	CAT	GTG	GAT	CTT	CTT	GCG	GGC	GCC	GCA	234
GTG	GTT	TGC	TCA	TCC	CTG	TAC	ATC	GGG	GAC	CTG	TGT	GGC	273
TCT	CTC	TTT	TTG	GCG	GGA	CAA	CTA	TTC	ACC	TTT	CAG	CCC	312
CGC	CGT	CAT	TGG	ACT	GTG	CAA	GAC	TGC	AAC	TGC	TCC	ATC	351
TAT	ACA	GGC	CAC	GTC	ACC	GGC	CAC	AGG	ATG	GCT	TGG	GAC	390
ATG	ATG	ATG	AAC	TGG	TCA	CCC	ACA	ACC	ACT	CTG	GTC	CTA	429
TCT	AGC	ATC	TTG	AGG	GTA	CCT	GAG	ATT	TGT	GCG	AGT	GTG	468
ATA	TTT	GGT	GGC	CAT	TGG	GGG	ATA	CTA	CTA	GCC	GTT	GCC	507
TAC	TTT	GGC	ATG	GCT	GGC	AAC	TGG	CTA	AAA	GTT	CTG	GCT	546
GTT	CTG	TTC	CTA	TTT	GCA	GGG	GTT	GAA	GCA				576

20

25

## (2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

30

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK7

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

①

20

(i)

18

25

**ORIGINAL SOURCE:**

(xi)

## 30

35

- 111 -

Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
                             80                            85                            90  
 Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg  
                             95                            100                            105  
 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
                             110                            115                            120  
 5 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                             125                            130                            135  
 Ser Pro Thr Ala Ala Leu Val Met Ala Gln Leu Leu Arg Ile Pro  
                             140                            145                            150  
 Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu  
                             155                            160                            165  
 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val  
                             170                            175                            180  
 10 Val Val Val Leu Leu Leu Phe Thr Gly Val Asp Ala  
                             185                            190

## (2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 20 (C) INDIVIDUAL ISOLATE: DR1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

His Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp  
                             5                            10                            15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu  
                             20                            25                            30  
 25 His Ala Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser  
                             35                            40                            45  
 Arg Cys Trp Val Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly  
                             50                            55                            60  
 Lys Leu Pro Thr Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val  
                             65                            70                            75  
 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
                             80                            85                            90  
 30 Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg  
                             95                            100                            105  
 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
                             110                            115                            120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                             125                            130                            135  
 Ser Pro Thr Thr Ala Leu Val Met Ala Gln Leu Leu Arg Ile Pro  
                             140                            145                            150  
 35

- 112 -

Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu  
 155 160  
 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Val Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala  
 185 190

5

## (2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DR4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

15 His Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp  
 5 10 15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Thr Ser  
 35 40 45  
 Arg Cys Trp Val Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly  
 50 55 60  
 20 Lys Leu Pro Thr Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val  
 65 70 75  
 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 His His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 25 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 30 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:56:

- 35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids

- 113 -

(B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S14

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

	Tyr	Gln	Val	Arg	Asn	Ser	Thr	Gly	Leu	Tyr	His	Val	Thr	Asn	Asp	
					5					10					15	
	Cys	Pro	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Thr	Ala	Asp	Ala	Ile	Leu	
					20					25					30	
10	His	Ala	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Gly	Asn	Thr	Ser	
					35					40					45	
	Arg	Cys	Trp	Val	Ala	Met	Thr	Pro	Thr	Val	Ala	Thr	Arg	Asp	Gly	
					50					55					60	
	Lys	Leu	Pro	Ala	Thr	Gln	Leu	Arg	Arg	Tyr	Ile	Asp	Leu	Leu	Val	
					65					70					75	
	Gly	Ser	Ala	Thr	Leu	Cys	Ser	Ala	Leu	Tyr	Val	Gly	Asp	Leu	Cys	
					80					85					90	
15	Gly	Ser	Val	Phe	Leu	Val	Gly	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	
					95					100					105	
	Arg	Leu	Trp	Thr	Thr	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly	
					110					115					120	
	His	Ile	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	
					125					130					135	
	Ser	Pro	Thr	Thr	Ala	Leu	Val	Val	Ala	Gln	Leu	Leu	Arg	Ile	Pro	
20					140					145					150	
	Gln	Ala	Ile	Leu	Asp	Met	Ile	Ala	Gly	Ala	His	Trp	Gly	Val	Leu	
					155					160					165	
	Ala	Gly	Ile	Ala	Tyr	Phe	Ser	Met	Val	Gly	Asn	Trp	Ala	Lys	Val	
					170					175					180	
	Leu	Val	Val	Leu	Leu	Leu	Phe	Ala	Gly	Val	Asp	Ala				
					185					190						

## (2) INFORMATION FOR SEQ ID NO:57: ..

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S18

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

- 114 -

Tyr Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp  
                                   5                                  10                                  15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Thr Ile Leu  
                                   20                                  25                                  30  
 His Ser Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser  
                                   35                                  40                                  45  
 Arg Cys Trp Val Pro Val Ala Pro Thr Val Ala Thr Arg Asp Gly  
 5                                  50                                  55                                  60  
 Lys Leu Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val  
                                   65                                  70                                  75  
 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
                                   80                                  85                                  90  
 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg  
                                   95                                  100                                  105  
 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 10                                  110                                  115                                  120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                                   125                                  130                                  135  
 Ser Pro Thr Thr Ala Leu Val Ile Ala Gln Leu Leu Arg Val Pro  
                                   140                                  145                                  150  
 Gln Ala Val Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu  
                                   155                                  160                                  165  
 Ala Gly Ile Ala Tyr Phe Ser Met Ala Gly Asn Trp Ala Lys Val  
 15                                  170                                  175                                  180  
 Leu Leu Val Leu Leu Leu Phe Ala Gly Val Asp Ala  
                                   185                                  190

## (2) INFORMATION FOR SEQ ID NO:58:

20 (i) SEQUENCE CHARACTERISTICS:  
           (A) LENGTH: 192 amino acids  
           (B) TYPE: amino acid  
           (C) STRANDEDNESS: unknown  
           (D) TOPOLOGY: unknown

25 (vi) ORIGINAL SOURCE:  
           (A) ORGANISM: homosapiens  
           (C) INDIVIDUAL ISOLATE: SW1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Tyr Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr Asn Asp  
                                   5                                  10                                  15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Ala Ile Leu  
 30                                  20                                  25                                  30  
 His Ser Pro Gly Cys Val Pro Cys Val Arg Glu Asp Gly Ala Pro  
                                   35                                  40                                  45  
 Lys Cys Trp Val Ala Val Ala Pro Thr Val Ala Thr Arg Asp Gly  
                                   50                                  55                                  60  
 Lys Leu Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val  
                                   65                                  70                                  75  
 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
 35                                  80                                  85                                  90



- 115 -

Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
                                   95                                  100                                  105  
 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
                                   110                                  115                                  120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                                   125                                  130                                  135  
 5 Ser Pro Thr Thr Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro  
                                   140                                  145                                  150  
 Gln Ala Val Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu  
                                   155                                  160                                  165  
 Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val  
                                   170                                  175                                  180  
 Leu Ile Val Leu Leu Leu Phe Ser Gly Val Asp Ala  
                                   185                                  190

## (2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US11

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Tyr Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr Asn Asp  
                                   5                                  10                                  15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile Leu  
                                   20                                  25                                  30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser  
                                   35                                  40                                  45  
 25 Arg Cys Trp Val Ala Met Thr Pro Thr Val Ala Thr Arg Asp Gly  
                                   50                                  55                                  60  
 Lys Leu Pro Thr Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val  
                                   65                                  70                                  75  
 Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
                                   80                                  85                                  90  
 Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg  
                                   95                                  100                                  105  
 30 Arg His Trp Thr Thr Gln Gly Cys Asn Cys Ser Ile Tyr Pro Gly  
                                   110                                  115                                  120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                                   125                                  130                                  135  
 Ser Pro Thr Ala Ala Leu Val Val Ala Gln Leu Leu Arg Ile Pro  
                                   140                                  145                                  150  
 Gln Ala Ile Leu Asp Met Ile Ala Gly Ala His Trp Gly Val Leu  
                                   155                                  160                                  165  
 35

- 116 -

Ala Gly Ile Ala Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: D1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met  
 15 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asp Asn Ser Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Gly  
 50 55 60  
 Asn Val Pro Thr Thr Ala Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Ile Ser Gln Leu Phe Thr Leu Ser Pro Arg  
 95 100 105  
 Arg His Glu Thr Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Val Met Asp Met Val Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

## (2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown

- 117 -

(D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: D3

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

	Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Val	Tyr	Gln	Val	Thr	Asn	Asp	
					5					10					15	
	Cys	Ser	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Thr	Ala	Asp	Met	Ile	Met	
					20					25					30	
	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Asp	Asn	Ser	Ser	
					35					40					45	
10	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ser	
					50					55					60	
	Ser	Val	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val	
					65					70					75	
	Gly	Ala	Ala	Ala	Phe	Cys	Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys	
					80					85					90	
	Gly	Ser	Val	Phe	Leu	Val	Ser	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	
					95					100					105	
15	Arg	His	Glu	Thr	Val	Gln	Glu	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly	
					110					115					120	
	His	Val	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	
					125					130					135	
	Ser	Pro	Thr	Ala	Ala	Leu	Val	Val	Ser	Gln	Leu	Leu	Arg	Ile	Pro	
					140					145					150	
	Gln	Ala	Val	Val	Asp	Met	Val	Ala	Gly	Ala	His	Trp	Gly	Val	Leu	
					155					160					165	
20	Ala	Gly	Leu	Ala	Tyr	Ser	Met	Val	Gly	Asn	Trp	Ala	Lys	Val		
					170					175					180	
	Leu	Ile	Val	Met	Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly				
					185					190						

25 (2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

30 (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: DK1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

35	Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Val	Tyr	His	Val	Thr	Asn	Asp	
					5					10					15	

- 118 -

Cys Ser Asn Ser Ser Il Val Tyr Glu Ala Val Asp Val Ile Met  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn His S r  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala  
 50 55 60  
 5 Ser Ile Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 Arg His Glu Thr Ala Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 10 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Val Leu Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Ala Gly Asn Trp Ala Lys Val  
 170 175 180  
 15 Leu Ile Val Leu Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

## (2) INFORMATION FOR SEQ ID NO:63:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 25 (C) INDIVIDUAL ISOLATE: HK3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Tyr Glu Val Arg Asn Val Ser Gly Ile Tyr His Val Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Ser Ser Val Val Tyr Glu Thr Ala Asp Met Ile Met  
 20 25 30  
 30 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Val  
 50 55 60  
 Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 35

Q

10

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 192 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

[illegible]

- 120 -

Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

## (2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 15 Cys Ser Asn Leu Ser Ile Val Tyr Glu Thr Thr Asp Met Ile Met  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Ala Pro Thr Leu Ala Ala Arg Asn Ala  
 50 55 60  
 Ser Val Pro Thr Thr Ala Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 20 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 25 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 30 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

## (2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown

- 121 -

(D) · · TOPOLOGY: · · unknown

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: HK8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

5	Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Ile	Tyr	His	Val	Thr	Asn	Asp
					5					10					15
	Cys	Ser	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Thr	Ala	Asp	Met	Ile	Met
					20					25					30
	His	Thr	Pro	Gly	Cys	Met	Pro	Cys	Val	Arg	Glu	Asn	Asn	Ser	Ser
					35					40					45
10	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Val
					50					55					60
	Ser	Val	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val
					65					70					75
	Gly	Ala	Ala	Ala	Phe	Cys	Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys
					80					85					90
	Gly	Ser	Val	Phe	Leu	Val	Ser	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg
					95					100					105
15	Arg	His	Glu	Thr	Val	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly
					110					115					120
	His	Val	Ser	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp
					125					130					135
	Ser	Pro	Thr	Thr	Ala	Leu	Val	Val	Ser	Gln	Leu	Leu	Arg	Ile	Pro
					140					145					150
	Gln	Ala	Ile	Val	Asp	Met	Val	Ala	Gly	Ala	His	Trp	Gly	Val	Leu
					155					160					165
20	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val	Gly	Asn	Trp	Ala	Lys	Val
					170					175					180
	Leu	Ile	Val	Met	Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly			
					185					190					

(2) INFORMATION FOR SEQ ID NO:67:

25 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 192 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

30 (vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: IND5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

35 Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp  
5 10 15  
Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met  
20 25 30

- 122 -

His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ser Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala  
 50 55 60  
 Ser Val Ser Thr Thr Thr Ile Arg His His Val Asp Leu Leu Val  
 65 70 75  
 5 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 10 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Ile Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190  
 15

## (2) INFORMATION FOR SEQ ID NO:68:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown  
 20

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: IND8

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Phe Ser  
 35 40 45  
 30 Ser Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala  
 50 55 60  
 Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 35



- 123 -

Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Ile Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

10 (2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

15 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: P10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 20 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Met Ile Met  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Asn Asn Ser Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ser  
 50 55 60  
 Ser Val Pro Thr Thr Ala Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 25 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Leu Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 Arg His Trp Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 30 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Ile Leu Asp Val Val Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 35 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

- 124 -

## (2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S9

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

10 Tyr Glu Val Arg Asn Val Ser Gly Ala Tyr His Val Thr Asn Asp  
                                   5                                  10                                  15  
   Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Val Ile Met  
                                   20                                  25                                  30  
   His Thr Pro Gly Cys Val Pro Cys Val Gln Glu Gly Asn Ser Ser  
                                   35                                  40                                  45  
 15 Gln Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala  
                                   50                                  55                                  60  
   Thr Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val  
                                   65                                  70                                  75  
   Gly Ala Ala Val Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
                                   80                                  85                                  90  
   Gly Ser Val Phe Leu Ile Ser Gln Leu Phe Thr Ile Ser Pro Arg  
                                   95                                  100                                  105  
 20 Arg His Glu Thr Val Gln Asn Cys Asn Cys Ser Ile Tyr Pro Gly  
                                   110                                  115                                  120  
   His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                                   125                                  130                                  135  
   Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
                                   140                                  145                                  150  
   Gln Ala Val Met Asp Met Val Ala Gly Ala His Trp Gly Val Leu  
                                   155                                  160                                  165  
 25 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
                                   170                                  175                                  180  
   Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
                                   185                                  190

## (2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens

- 125 -

## (C) INDIVIDUAL ISOLATE: S45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

	Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Ala	Tyr	His	Val	Thr	Asn	Asp	
					5					10					15	
5	Cys	Ser	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Val	Asp	Val	Ile	Leu	
					20					25					30	
	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Asn	Asn	Ser	Ser	
					35					40					45	
	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ser	
					50					55					60	
	Ser	Val	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val	
					65					70					75	
10	Gly	Ala	Ala	Ala	Phe	Cys	Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys	
					80					85					90	
	Gly	Ser	Val	Phe	Leu	Val	Ser	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	
					95					100					105	
	Arg	His	Glu	Thr	Val	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly	
					110					115					120	
	His	Val	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	
					125					130					135	
15	Ser	Pro	Thr	Ala	Ala	Leu	Val	Val	Ser	Gln	Leu	Leu	Arg	Ile	Pro	
					140					145					150	
	Gln	Ala	Val	Val	Asp	Met	Val	Ala	Gly	Ala	His	Trp	Gly	Val	Leu	
					155					160					165	
	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val	Gly	Asn	Trp	Ala	Lys	Val	
					170					175					180	
20	Leu	Ile	Val	Met	Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly				
					185					190						

## (2) INFORMATION FOR SEQ ID NO:72:

## (i) SEQUENCE CHARACTERISTICS:

	(A)	LENGTH:	192 amino acids
	(B)	TYPE:	amino acid
25	(C)	STRANDEDNESS:	unknown
	(D)	TOPOLOGY:	unknown

## (vi) ORIGINAL SOURCE:

(A)	ORGANISM:	homosapiens
(C)	INDIVIDUAL ISOLATE:	SA10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

30	Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Met	Tyr	His	Val	Thr	Asn	Asp	
					5					10					15	
	Cys	Ser	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Ala	Asp	Met	Ile	Met	
					20					25					30	
	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Asn	Asn	Ser	Ser	
					35					40					45	
35	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ser	
					50					55					60	

- 126 -

Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 5 Arg Tyr Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 Arg Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Ile Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 10 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

## (2) INFORMATION FOR SEQ ID NO:73:

15

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

20

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 25 Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Ala Asn Ser Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Thr  
 50 55 60  
 30 Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Ala Phe Cys Ser Val Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 35 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135

- 127 -

Ser Pro Thr Ala Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 5 Leu Ile Val Met Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

## (2) INFORMATION FOR SEQ ID NO:74:

## (i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

15

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T3

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Tyr Glu Val Arg Asn Val Ser Gly Val Tyr Tyr Val Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Ser Ser Ile Val Tyr Glu Thr Ala Asp Met Ile Met  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Ser Asn Ser Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala Arg Asn Ala  
 50 55 60  
 Ser Val Pro Thr Lys Thr Ile Arg Arg His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val Gly Asp Leu Cys  
 80 85 90  
 25 Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Phe Ser Pro Arg  
 95 100 105  
 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro  
 140 145 150  
 30 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val  
 170 175 180  
 Leu Ile Val Leu Leu Leu Phe Ala Gly Val Asp Gly  
 185 190

35

- 128 -

## (2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

10	Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Met	Tyr	His	Val	Thr	Asn	Asp	5	10	15
	Cys	Ser	Asn	Ser	Ser	Ile	Val	Phe	Glu	Ala	Ala	Asp	Leu	Ile	Met	20	25	30
	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Gly	Asn	Ser	Ser	35	40	45
	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Thr	50	55	60
15	Ser	Val	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val	65	70	75
	Gly	Ala	Ala	Ala	Phe	Cys	Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys	80	85	90
	Gly	Ser	Val	Phe	Leu	Val	Ser	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg	95	100	105
	Arg	His	Glu	Thr	Leu	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly	110	115	120
20	His	Leu	Ser	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	125	130	135
	Ser	Pro	Thr	Thr	Ala	Leu	Val	Val	Ser	Gln	Leu	Leu	Arg	Ile	Pro	140	145	150
	Gln	Ala	Val	Met	Asp	Met	Val	Thr	Gly	Ala	His	Trp	Gly	Val	Leu	155	160	165
	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Ala	Gly	Asn	Trp	Ala	Lys	Val	170	175	180
25	Leu	Ile	Val	Met	Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly				185	190	

## (2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US6

- 129 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

	Tyr	Glu	Val	Arg	Asn	Val	Ser	Gly	Met	Tyr	His	Val	Thr	Asn	Asp
					5					10					15
	Cys	Ser	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Ala	Asp	Met	Ile	Met
					20					25					30
5	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Asn	Asn	Ser	Ser
					35					40					45
	Arg	Cys	Trp	Val	Ala	Leu	Thr	Pro	Thr	Leu	Ala	Ala	Arg	Asn	Ala
					50					55					60
	Ser	Val	Pro	Thr	Thr	Thr	Ile	Arg	Arg	His	Val	Asp	Leu	Leu	Val
					65					70					75
	Gly	Ala	Ala	Thr	Phe	Cys	Ser	Ala	Met	Tyr	Val	Gly	Asp	Leu	Cys
					80					85					90
10	Gly	Ser	Val	Phe	Leu	Ile	Ser	Gln	Leu	Phe	Thr	Phe	Ser	Pro	Arg
					95					100					105
	Gln	His	Glu	Thr	Val	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly
					110					115					120
	His	Val	Ser	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp
					125					130					135
	Ser	Pro	Thr	Ala	Ala	Leu	Val	Val	Ser	Gln	Leu	Leu	Arg	Ile	Pro
					140					145					150
15	Gln	Ala	Val	Met	Asp	Met	Val	Ala	Gly	Ala	His	Trp	Gly	Val	Leu
					155					160					165
	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Val	Gly	Asn	Trp	Ala	Lys	Val
					170					175					180
	Leu	Ile	Val	Leu	Leu	Leu	Phe	Ala	Gly	Val	Asp	Gly			
					185					190					

20

## (2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 192 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

25

- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: T2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

30	Ala	Gln	Val	Arg	Asn	Thr	Ser	Arg	Gly	Tyr	Met	Val	Thr	Asn	Asp
					5					10					15
	Cys	Ser	Asn	Glu	Ser	Ile	Thr	Trp	Gln	Leu	Gln	Ala	Ala	Val	Leu
					20					25					30
	His	Val	Pro	Gly	Cys	Ile	Pro	Cys	Glu	Arg	Leu	Gly	Asn	Thr	Ser
					35					40					45
	Arg	Cys	Trp	Ile	Pro	Val	Thr	Pro	Asn	Val	Ala	Val	Arg	Gln	Pro
					50					55					60

35

- 130 -

Gly Ala Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val  
                   65                  70                  75  
 Met Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
                   80                  85                  90  
 Gly Gly Val Met Leu Ala Ala Gln Met Phe Ile Val Ser Pro Arg  
                   95                  100                  105  
 Arg His Trp Phe Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly  
 5                   110                  115                  120  
 Thr Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                   125                  130                  135  
 Ser Pro Thr Ala Thr Met Ile Leu Ala Tyr Ala Met Arg Val Pro  
                   140                  145                  150  
 Glu Val Ile Ile Asp Ile Ile Gly Gly Ala His Trp Gly Val Met  
                   155                  160                  165  
 10 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val  
                   170                  175                  180  
 Ile Val Ile Leu Leu Leu Ala Ala Gly Val Asp Ala  
                   185                  190

## (2) INFORMATION FOR SEQ ID NO:78:

15 (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 192 amino acids  
       (B) TYPE: amino acid  
       (C) STRANDEDNESS: unknown  
       (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 20 (A) ORGANISM: homosapiens  
       (C) INDIVIDUAL ISOLATE: T4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Ala Gln Val Lys Asn Thr Thr Asn Ser Tyr Met Val Thr Asn Asp  
                   5                  10                  15  
 Cys Ser Asn Asp Ser Ile Thr Trp Gln Leu Gln Ala Ala Val Leu  
 25                   20                  25                  30  
 His Val Pro Gly Cys Val Pro Cys Glu Lys Thr Gly Asn Thr Ser  
                   35                  40                  45  
 Arg Cys Trp Ile Pro Val Ser Pro Asn Val Ala Val Arg Gln Pro  
                   50                  55                  60  
 Gly Ala Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val  
                   65                  70                  75  
 Met Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
 30                   80                  85                  90  
 Gly Gly Val Met Leu Ala Ala Gln Met Phe Ile Val Ser Pro Gln  
                   95                  100                  105  
 His His Trp Phe Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
                   110                  115                  120  
 Thr Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                   125                  130                  135  
 35 Ser Pro Thr Ala Thr Met Ile Leu Ala Tyr Ala Met Arg Val Pro  
                   140                  145                  150



- 131 -

°

Glu Val Ile Leu Asp Il Val Ser Gly Ala His Trp Gly Val Met  
 155 160 165  
 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val  
 170 175 180  
 Val Val Ile Leu Leu Leu Ala Ala Gly Val Asp Ala  
 185 190

5

## (2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

15

Ala Glu Val Lys Asn Thr Ser Thr Ser Tyr Met Val Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Asp Ser Ile Thr Trp Gln Leu Gln Ala Ala Val Leu  
 20 25 30  
 His Val Pro Gly Cys Val Pro Cys Glu Arg Val Gly Asn Ala Ser  
 35 40 45  
 Arg Cys Trp Ile Pro Val Ser Pro Asn Val Ala Val Gln Arg Pro  
 50 55 60  
 Gly Ala Leu Thr Gln Gly Leu Arg Thr His Ile Asp Met Val Val  
 65 70 75  
 Met Ser Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Gly Val Met Leu Ala Ala Gln Met Phe Ile Ile Ser Pro Gln  
 95 100 105  
 His His Trp Phe Val Gln Glu Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 Thr Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Thr Met Ile Leu Ala Tyr Ala Met Arg Val Pro  
 140 145 150  
 Glu Val Ile Ile Asp Ile Ile Ser Gly Ala His Trp Gly Val Met  
 155 160 165  
 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val  
 170 175 180  
 Val Val Ile Leu Leu Leu Thr Ala Gly Val Asp Ala  
 185 190

30

## (2) INFORMATION FOR SEQ ID NO:80:

35

(i) SEQUENCE CHARACTERISTICS:

- 132 -

- (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

	Val	Gln	Val	Lys	Asn	Thr	Ser	Thr	Ser	Tyr	Met	Val	Thr	Asn	Asp	
					5					10						15
10	Cys	Ser	Asn	Asp	Ser	Ile	Thr	Trp	Gln	Leu	Glu	Ala	Ala	Val	Leu	
					20					25						30
	His	Val	Pro	Gly	Cys	Val	Pro	Cys	Glu	Lys	Val	Gly	Asn	Thr	Ser	
					35					40						45
	Arg	Cys	Trp	Ile	Pro	Val	Ser	Pro	Asn	Val	Ala	Val	Gln	Arg	Pro	
					50					55						60
	Gly	Ala	Leu	Thr	Gln	Gly	Leu	Arg	Thr	His	Ile	Asp	Met	Val	Val	
					65					70						75
15	Met	Ser	Ala	Thr	Leu	Cys	Ser	Ala	Leu	Tyr	Val	Gly	Asp	Phe	Cys	
					80					85						90
	Gly	Gly	Met	Met	Leu	Ala	Ala	Gln	Met	Phe	Ile	Val	Ser	Pro	Arg	
					95					100						105
	His	His	Ser	Phe	Val	Gln	Glu	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly	
					110					115						120
	Thr	Ile	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	
					125					130						135
20	Ser	Pro	Thr	Ala	Thr	Leu	Ile	Leu	Ala	Tyr	Val	Met	Arg	Val	Pro	
					140					145						150
	Glu	Val	Ile	Ile	Asp	Ile	Ile	Ser	Gly	Ala	His	Trp	Gly	Val	Leu	
					155					160						165
	Phe	Gly	Leu	Ala	Tyr	Phe	Ser	Met	Gln	Gly	Ala	Trp	Ala	Lys	Val	
					170					175						180
	Val	Val	Ile	Leu	Leu	Leu	Ala	Ala	Gly	Val	Asp	Ala				
25					185					190						

(2) INFORMATION FOR SEQ ID NO:81:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK8

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

- 133 -

Val Glu Val Arg Asn Ile Ser Ser Ser Tyr Tyr Ala Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Asn Ser Ile Thr Trp Gln Leu Thr Asp Ala Val Leu  
 20 25 30  
 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu  
 35 40 45  
 5 Arg Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg  
 50 55 60  
 Gly Ala Leu Thr His Asn Leu Arg Thr His Val Asp Val Ile Val  
 65 70 75  
 Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Val Cys  
 80 85 90  
 Gly Ala Val Met Ile Val Ser Gln Ala Leu Ile Ile Ser Pro Glu  
 95 100 105  
 10 Arg His Asn Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly  
 110 115 120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp  
 125 130 135  
 Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro  
 140 145 150  
 Glu Leu Ala Leu Gln Val Val Phe Gly Gly His Trp Gly Val Val  
 155 160 165  
 15 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val  
 170 175 180  
 Ile Ala Ile Leu Leu Leu Val Ala Gly Val Asp Ala  
 185 190

20 (2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

25 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Val Glu Val Arg Asn Thr Ser Ser Ser Tyr Tyr Ala Thr Asn Asp  
 5 10 15  
 30 Cys Ser Asn Asn Ser Ile Thr Trp Gln Leu Thr Asn Ala Val Leu  
 20 25 30  
 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu  
 35 40 45  
 His Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg  
 50 55 60  
 Gly Ala Leu Thr His Asn Leu Arg Ala His Ile Asp Met Ile Val  
 65 70 75  
 35

- 134 -

Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Val Cys  
 80 85 90  
 Gly Ala Val Met Ile Val Ser Gln Ala Phe Ile Val Ser Pro Glu  
 95 100 105  
 His His His Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly  
 110 115 120  
 5 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp  
 125 130 135  
 Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro  
 140 145 150  
 Glu Leu Val Leu Glu Val Val Phe Gly Gly His Trp Gly Val Val  
 155 160 165  
 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val  
 170 175 180  
 10 Ile Ala Ile Leu Leu Leu Val Ala Gly Val Asp Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 15 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW3

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Val Glu Val Arg Asn Ile Ser Ser Ser Tyr Tyr Ala Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Ser Ser Ile Thr Trp Gln Leu Thr Asn Ala Val Leu  
 20 25 30  
 25 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu  
 35 40 45  
 His Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg  
 50 55 60  
 Gly Ala Leu Thr His Asn Leu Arg Ala His Val Asp Met Ile Val  
 65 70 75  
 Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Met Cys  
 80 85 90  
 30 Gly Ala Val Met Ile Val Ser Gln Ala Phe Ile Ile Ser Pro Glu  
 95 100 105  
 Arg His Asn Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly  
 110 115 120  
 Arg Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp  
 125 130 135  
 Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro  
 140 145 150  
 35 Glu Leu Val Leu Glu Val Val Phe Gly Gly His Trp Gly Val Val  
 155 160 165

- 135 -

Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val  
 170 175 180  
 Ile Ala Ile Leu Leu Leu Val Ala Gly Val Asp Ala  
 185 190

## 5 (2) INFORMATION FOR SEQ ID NO:84:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- 10 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

15 Val Glu Val Arg Asn Thr Ser Phe Ser Tyr Tyr Ala Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Asn Ser Ile Thr Trp Gln Leu Thr Asn Ala Val Leu  
 20 25 30  
 His Leu Pro Gly Cys Val Pro Cys Glu Asn Asp Asn Gly Thr Leu  
 35 40 45  
 Arg Cys Trp Ile Gln Val Thr Pro Asn Val Ala Val Lys His Arg  
 50 55 60  
 20 Gly Ala Leu Thr His Asn Leu Arg Thr His Val Asp Val Ile Val  
 65 70 75  
 Met Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Val Cys  
 80 85 90  
 Gly Ala Val Met Ile Ala Ser Gln Ala Phe Ile Ile Ser Pro Glu  
 95 100 105  
 Arg His Asn Phe Thr Gln Glu Cys Asn Cys Ser Ile Tyr Gln Gly  
 110 115 120  
 25 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Leu Asn Trp  
 125 130 135  
 Ser Pro Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro  
 140 145 150  
 Glu Leu Val Leu Glu Val Val Phe Gly Gly His Trp Gly Val Val  
 155 160 165  
 Phe Gly Leu Ala Tyr Phe Ser Met Gln Gly Ala Trp Ala Lys Val  
 170 175 180  
 30 Ile Ala Ile Leu Leu Leu Val Ala Gly Val Asp Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:85:

- 35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids

- 136 -

(B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S83

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

	Val	Glu	Val	Lys	Asp	Thr	Gly	Asp	Ser	Tyr	Met	Pro	Thr	Asn	Asp	
					5					10						15
	Cys	Ser	Asn	Ser	Ser	Ile	Val	Trp	Gln	Leu	Glu	Gly	Ala	Val	Leu	
10					20					25						30
	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Glu	Arg	Thr	Ala	Asn	Val	Ser	
					35					40						45
	Arg	Cys	Trp	Val	Pro	Val	Ala	Pro	Asn	Leu	Ala	Ile	Ser	Gln	Pro	
					50					55						60
	Gly	Ala	Leu	Thr	Lys	Gly	Leu	Arg	Ala	His	Ile	Asp	Ile	Ile	Val	
					65					70						75
	Met	Ser	Ala	Thr	Val	Cys	Ser	Ala	Leu	Tyr	Val	Gly	Asp	Val	Cys	
15					80					85						90
	Gly	Ala	Leu	Met	Leu	Ala	Ala	Gln	Val	Val	Val	Val	Ser	Pro	Gln	
					95					100						105
	His	His	Thr	Phe	Val	Gln	Glu	Cys	Asn	Cys	Ser	Ile	Tyr	Pro	Gly	
					110					115						120
	Arg	Ile	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	
					125					130						135
20	Ser	Pro	Thr	Thr	Thr	Met	Leu	Leu	Ala	Tyr	Leu	Val	Arg	Ile	Pro	
					140					145						150
	Glu	Val	Ile	Leu	Asp	Ile	Val	Thr	Gly	Gly	His	Trp	Gly	Val	Met	
					155					160						165
	Phe	Gly	Leu	Ala	Tyr	Phe	Ser	Met	Gln	Gly	Ser	Trp	Ala	Lys	Val	
					170					175						180
	Ile	Val	Ile	Leu	Leu	Leu	Thr	Ala	Gly	Val	Glu	Ala				
					185					190						

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

- 137 -

0 Leu Glu Trp Arg Asn Val Ser Gly Leu Tyr Val Leu Thr Asn Asp  
     5                    10                    15  
 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu  
     20                    25                    30  
 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser  
     35                    40                    45  
 5 Thr Cys Trp Thr Ser Val Thr Pro Thr Val Ala Val Arg Tyr Val  
     50                    55                    60  
 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val  
     65                    70                    75  
 Gly Ala Ala Thr Met Cys Ser Ala Leu Tyr Val Gly Asp Val Cys  
     80                    85                    90  
 Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg  
     95                    100                    105  
 10 Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly  
     110                    115                    120  
 His Leu Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
     125                    130                    135  
 Ser Pro Ala Val Gly Met Val Val Ala His Val Leu Arg Leu Pro  
     140                    145                    150  
 Gln Thr Leu Phe Asp Ile Ile Ala Gly Ala His Trp Gly Ile Met  
     155                    160                    165  
 15 Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val  
     170                    175                    180  
 Ala Ile Ile Met Val Met Phe Ser Gly Val Asp Ala  
     185                    190

## (2) INFORMATION FOR SEQ ID NO:87:

20

(i)

## SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi)

## ORIGINAL SOURCE:

25

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK10

(xi)

## SEQUENCE DESCRIPTION: SEQ ID NO:87:

30 Leu Glu Trp Arg Asn Val Ser Gly Leu Tyr Val Leu Thr Asn Asp  
     5                    10                    15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu  
     20                    25                    30  
 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser  
     35                    40                    45  
 Thr Cys Trp Thr Ser Val Thr Pro Thr Val Ala Val Arg Tyr Val  
     50                    55                    60  
 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val  
     65                    70                    75  
 35 Gly Ala Ala Thr Met Cys Ser Ala Leu Tyr Val Gly Asp Met Cys  
     80                    85                    90

- 138 -

Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg  
 95 100 105  
 Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly  
 110 115 120  
 His Leu Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 5 Ser Pro Ala Val Gly Met Val Val Ala His Val Leu Arg Leu Pro  
 140 145 150  
 Gln Thr Leu Phe Asp Ile Ile Ala Gly Ala His Trp Gly Ile Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val  
 170 175 180  
 Ala Ile Ile Met Val Met Phe Ser Gly Val Asp Ala  
 185 190

10

## (2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

15

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

20

Leu Glu Trp Arg Asn Thr Ser Gly Leu Tyr Val Leu Thr Asn Asp  
 5 10 15  
 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser  
 35 40 45  
 25 Thr Cys Trp Thr Pro Val Thr Pro Thr Val Ala Val Arg Tyr Val  
 50 55 60  
 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val  
 65 70 75  
 Gly Ala Ala Thr Met Cys Ser Ala Leu Tyr Val Gly Asp Met Cys  
 80 85 90  
 Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg  
 95 100 105  
 30 Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly  
 110 115 120  
 His Leu Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Ala Val Gly Met Val Val Ala His Val Leu Arg Leu Pro  
 140 145 150  
 Gln Thr Val Phe Asp Ile Ile Ala Gly Ala His Trp Gly Ile Leu  
 155 160 165  
 35



- 139 -

Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val  
 170 175 180  
 Ala Ile Ile Met Val Met Phe Ser Gly Val Asp Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S52

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Leu Glu Trp Arg Asn Thr Ser Gly Leu Tyr Val Leu Thr Asn Asp  
 5 10 15  
 15 Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala Asp Asp Val Ile Leu  
 20 25 30  
 His Thr Pro Gly Cys Val Pro Cys Val Gln Asp Gly Asn Thr Ser  
 35 40 45  
 Met Cys Trp Thr Pro Val Thr Pro Thr Val Ala Val Arg Tyr Val  
 50 55 60  
 Gly Ala Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val  
 65 70 75  
 20 Gly Ala Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Met Cys  
 80 85 90  
 Gly Ala Val Phe Leu Val Gly Gln Ala Phe Thr Phe Arg Pro Arg  
 95 100 105  
 Arg His Gln Thr Val Gln Thr Cys Asn Cys Ser Leu Tyr Pro Gly  
 110 115 120  
 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 25 Ser Pro Ala Val Gly Met Val Val Ala His Ile Leu Arg Leu Pro  
 140 145 150  
 Gln Thr Leu Phe Asp Ile Leu Ala Gly Ala His Trp Gly Ile Leu  
 155 160 165  
 Ala Gly Leu Ala Tyr Tyr Ser Met Gln Gly Asn Trp Ala Lys Val  
 170 175 180  
 30 Ala Ile Val Met Ile Met Phe Ser Gly Val Asp Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown

- 140 -

(D) TOPOLOGY: unknown

(vi)

ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S54

5

(xi)

SEQUENCE DESCRIPTION: SEQ ID NO:90:

	Leu	Glu	Trp	Arg	Asn	Thr	Ser	Gly	Leu	Tyr	Ile	Leu	Thr	Asn	Asp	
					5					10						15
	Cys	Ser	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Asp	Asp	Val	Ile	Leu	
					20					25						30
	His	Thr	Pro	Gly	Cys	Val	Pro	Cys	Val	Gln	Asp	Gly	Asn	Thr	Ser	
					35					40						45
10	Thr	Cys	Trp	Thr	Pro	Val	Thr	Pro	Thr	Val	Ala	Val	Arg	Tyr	Val	
					50					55						60
	Gly	Ala	Thr	Thr	Ala	Ser	Ile	Arg	Ser	His	Val	Asp	Leu	Leu	Val	
					65					70						75
	Gly	Ala	Ala	Thr	Leu	Cys	Ser	Ala	Leu	Tyr	Val	Gly	Asp	Met	Cys	
					80					85						90
	Gly	Ala	Val	Phe	Leu	Val	Gly	Gln	Ala	Phe	Thr	Phe	Arg	Pro	Arg	
					95					100						105
15	Arg	His	Gln	Thr	Val	Gln	Thr	Cys	Asn	Cys	Ser	Leu	Tyr	Pro	Gly	
					110					115						120
	His	Leu	Ser	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	
					125					130						135
	Ser	Pro	Ala	Val	Gly	Met	Val	Val	Ala	His	Ile	Leu	Arg	Leu	Pro	
					140					145						150
	Gln	Thr	Leu	Phe	Asp	Ile	Leu	Ala	Gly	Ala	His	Trp	Gly	Ile	Leu	
20					155					160						165
	Ala	Gly	Leu	Ala	Tyr	Tyr	Ser	Met	Gln	Gly	Asn	Trp	Ala	Lys	Val	
					170					175						180
	Ala	Ile	Ile	Met	Ile	Met	Phe	Ser	Gly	Val	Asp	Ala				
					185					190						

25

(2) INFORMATION FOR SEQ ID NO:91:

(i)

SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

30

(vi)

ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: Z4

(xi)

SEQUENCE DESCRIPTION: SEQ ID NO:91:

35

Glu	His	Tyr	Arg	Asn	Ala	Ser	Gly	Ile	Tyr	His	Ile	Thr	Asn	Asp
				5						10				15

- 141 -

0 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp His His Ile Leu  
     20 25 30  
 His Leu Pro Gly Cys Val Pro Cys Val Met Thr Gly Asn Thr Ser  
     35 40 45  
 Arg Cys Trp Thr Pro Val Thr Pro Thr Val Ala Val Ala His Pro  
     50 55 60  
 5 Gly Ala Pro Leu Glu Ser Phe Arg Arg His Val Asp Leu Met Val  
     65 70 75  
 Gly Ala Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
     80 85 90  
 Gly Gly Ala Phe Leu Met Gly Gln Met Ile Thr Phe Arg Pro Arg  
     95 100 105  
 Arg His Trp Thr Thr Gln Glu Cys Asn Cys Ser Ile Tyr Thr Gly  
     110 115 120  
 10 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
     125 130 135  
 Ser Pro Thr Thr Thr Leu Leu Leu Ala Gln Ile Met Arg Val Pro  
     140 145 150  
 Thr Ala Phe Leu Asp Met Val Ala Gly Gly His Trp Gly Val Leu  
     155 160 165  
 Ala Gly Leu Ala Tyr Phe Ser Met Gln Gly Asn Trp Ala Lys Val  
     170 175 180  
 15 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala  
     185 190

## (2) INFORMATION FOR SEQ ID NO:92:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown  
  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z1  
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

Val His Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp  
     5 10 15  
 Cys Pro Asn Thr Ser Ile Val Tyr Glu Thr Glu His His Ile Met  
     20 25 30  
 30 His Leu Pro Gly Cys Val Pro Cys Val Arg Thr Glu Asn Thr Ser  
     35 40 45  
 Arg Cys Trp Val Pro Leu Thr Pro Thr Val Ala Ala Pro Tyr Pro  
     50 55 60  
 Asn Ala Pro Leu Glu Ser Met Arg Arg His Val Asp Leu Met Val  
     65 70 75  
 Gly Ala Ala Thr Met Cys Ser Ala Phe Tyr Ile Gly Asp Leu Cys  
     80 85 90  
 35 Gly Gly Val Phe Leu Val Gly Gln Leu Phe Asp Phe Arg Pro Arg  
     95 100 105

- 142 -

Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly  
 110 115 120  
 His Val Ser Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Ser Ala Leu Ile Met Ala Gln Ile Leu Arg Ile Pro  
 140 145 150  
 5 Ser Ile Leu Gly Asp Leu Leu Thr Gly Gly His Trp Gly Val Leu  
 155 160 165  
 Ala Gly Leu Ala Phe Phe Ser Met Gln Ser Asn Trp Ala Lys Val  
 170 175 180  
 Ile Leu Val Leu Phe Leu Phe Ala Gly Val Glu Gly  
 185 190

10

## (2) INFORMATION FOR SEQ ID NO:93:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

15

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z6

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

20 Val Asn Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Glu His Gln Ile Leu  
 20 25 30  
 His Leu Pro Gly Cys Leu Pro Cys Val Arg Val Gly Asn Gln Ser  
 35 40 45  
 Arg Cys Trp Val Ala Leu Thr Pro Thr Val Ala Val Ser Tyr Ile  
 50 55 60  
 25 Gly Ala Pro Leu Asp Ser Leu Arg Arg His Val Asp Leu Met Val  
 65 70 75  
 Gly Ala Ala Thr Val Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys  
 80 85 90  
 Gly Gly Ala Phe Leu Val Gly Gln Met Phe Ser Phe Gln Pro Arg  
 95 100 105  
 Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Ala Gly  
 110 115 120  
 30 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Thr Leu Leu Leu Ala Gln Val Met Arg Ile Pro  
 140 145 150  
 Ser Thr Leu Val Asp Leu Leu Ala Gly Gly His Trp Gly Val Leu  
 155 160 165  
 Val Gly Leu Ala Tyr Phe Ser Met Gln Ala Asn Trp Ala Lys Val  
 170 175 180  
 35

- 143 -

Ile Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala  
185 190

## (2) INFORMATION FOR SEQ ID NO:94:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 192 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: Z7

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Val Asn Tyr His Asn Ala Ser Gly Val Tyr His Ile Thr Asn Asp  
5 10 15  
Cys Pro Asn Ser Ser Ile Met Tyr Glu Ala Glu His His Ile Leu  
20 25 30  
His Leu Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Gln Ser  
15 35 40 45  
Arg Cys Trp Val Ala Leu Thr Pro Thr Val Ala Ala Pro Tyr Ile  
50 55 60  
Gly Ala Pro Leu Glu Ser Ile Arg Arg His Val Asp Leu Met Val  
65 70 75  
Gly Ala Ala Thr Val Cys Ser Ala Leu Tyr Ile Gly Asp Leu Cys  
80 85 90  
20 Gly Gly Val Phe Leu Val Gly Gln Met Phe Ser Phe Gln Pro Arg  
95 100 105  
Arg His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile Tyr Ala Gly  
110 115 120  
His Val Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
125 130 135  
Ser Pro Thr Thr Thr Leu Val Leu Ala Gln Val Met Arg Ile Pro  
140 145 150  
25 Ser Thr Leu Val Asp Leu Leu Thr Gly Gly His Trp Gly Ile Leu  
155 160 165  
Ile Gly Val Ala Tyr Phe Cys Met Gln Ala Asn Trp Ala Lys Val  
170 175 180  
Ile Leu Val Leu Phe Leu Tyr Ala Gly Val Asp Ala  
185 190

30

## (2) INFORMATION FOR SEQ ID NO:95:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

35

## (vi) ORIGINAL SOURCE:

- 144 -

(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: DK13

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

5	Tyr	Asn	Tyr	Arg	Asn	Ser	Ser	Gly	Val	Tyr	His	Val	Thr	Asn	Asp	5	10	15
	Cys	Pro	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Thr	Asp	Tyr	His	Ile	Leu	20	25	30
	His	Leu	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Glu	Gly	Asn	Lys	Ser	35	40	45
	Thr	Cys	Trp	Val	Ser	Leu	Thr	Pro	Thr	Val	Ala	Ala	Gln	His	Leu	50	55	60
10	Asn	Ala	Pro	Leu	Glu	Ser	Leu	Arg	Arg	His	Val	Asp	Leu	Met	Val	65	70	75
	Gly	Gly	Ala	Thr	Leu	Cys	Ser	Ala	Leu	Tyr	Ile	Gly	Asp	Val	Cys	80	85	90
	Gly	Gly	Val	Phe	Leu	Val	Gly	Gln	Leu	Phe	Thr	Phe	Gln	Pro	Arg	95	100	105
	Arg	His	Trp	Thr	Thr	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Thr	Gly	110	115	120
15	His	Ile	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	125	130	135
	Ser	Pro	Thr	Ala	Thr	Leu	Val	Leu	Ala	Gln	Leu	Met	Arg	Ile	Pro	140	145	150
	Gly	Ala	Met	Val	Asp	Leu	Leu	Ala	Gly	Gly	His	Trp	Gly	Ile	Leu	155	160	165
20	Val	Gly	Ile	Ala	Tyr	Phe	Ser	Met	Gln	Ala	Asn	Trp	Ala	Lys	Val	170	175	180
	Ile	Leu	Val	Leu	Phe	Leu	Phe	Ala	Gly	Val	Asp	Ala				185	190	

## (2) INFORMATION FOR SEQ ID NO:96:

## (i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 192 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

30 (A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: SA1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

35	Val	Pro	Tyr	Arg	Asn	Ala	Ser	Gly	Val	Tyr	His	Val	Thr	Asn	Asp	5	10	15
	Cys	Pro	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Asp	Ser	Leu	Ile	Leu	20	25	30
	His	Ala	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Gln	Asp	Asn	Val	Ser	35	40	45

- 145 -

Arg Cys Trp Val Gln Ile Thr Pro Thr Leu Ser Ala Pro Thr Phe  
 50 55 60  
 Gly Ala Val Thr Ala Pro Leu Arg Arg Ala Val Asp Tyr Leu Ala  
 65 70 75  
 Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys  
 80 85 90  
 5 Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Arg Pro Arg  
 95 100 105  
 Gln His Thr Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly  
 110 115 120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Leu Met Ala Gln Met Leu Arg Ile Pro  
 140 145 150  
 10 Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu  
 155 160 165  
 Phe Ala Ala Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val  
 170 175 180  
 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Gly  
 185 190

15

## (2) INFORMATION FOR SEQ ID NO:97:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

20

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: *homo sapiens*  
 (C) INDIVIDUAL ISOLATE: SA4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

25 Val Pro Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asn Leu Ile Leu  
 20 25 30  
 His Ala Pro Gly Cys Val Pro Cys Val Arg Gln Asp Asn Val Ser  
 35 40 45  
 Lys Cys Trp Val Gln Ile Thr Pro Thr Leu Ser Ala Pro Asn Leu  
 50 55 60  
 30 Gly Ala Val Thr Ala Pro Leu Arg Arg Ala Val Asp Tyr Leu Ala  
 65 70 75  
 Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys  
 80 85 90  
 Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Arg Pro Arg  
 95 100 105  
 Gln His Thr Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly  
 110 115 120  
 35

- 146 -

His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Leu Met Ala Gln Leu Leu Arg Ile Pro  
 140 145 150  
 Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu  
 155 160 165  
 Phe Ala Ala Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val  
 170 175 180  
 Ile Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SAS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Val Pro Tyr Arg Asn Ala Ser Gly Val Tyr His Val Thr Asn Asp  
 5 10 15  
 Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp Asn Leu Ile Leu  
 20 25 30  
 His Ala Pro Gly Cys Val Pro Cys Val Lys Glu Gly Asn Val Ser  
 35 40 45  
 Arg Cys Trp Val Gln Ile Thr Pro Thr Leu Ser Ala Pro Asn Leu  
 50 55 60  
 Gly Ala Val Thr Ala Pro Leu Arg Arg Val Val Asp Tyr Leu Ala  
 65 70 75  
 Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys  
 80 85 90  
 Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Arg Pro Arg  
 95 100 105  
 Gln His Thr Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly  
 110 115 120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
 125 130 135  
 Ser Pro Thr Thr Ala Leu Val Met Ala Gln Val Leu Arg Ile Pro  
 140 145 150  
 Gln Val Val Ile Asp Ile Ile Ala Gly Gly His Trp Gly Val Leu  
 155 160 165  
 Phe Ala Val Ala Tyr Phe Ala Ser Ala Ala Asn Trp Ala Lys Val  
 170 175 180  
 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Gly  
 185 190

35



- 147 -

## (2) INFORMATION FOR SEQ ID NO:99:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA6

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

10	Val	Pro	Tyr	Arg	Asn	Ala	Ser	Gly	Val	Tyr	His	Val	Thr	Asn	Asp
					5					10					15
	Cys	Pro	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Asp	Asp	Leu	Ile	Leu
					20					25					30
	His	Ala	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Lys	Asp	Asn	Val	Ser
					35					40					45
	Arg	Cys	Trp	Val	His	Ile	Thr	Pro	Thr	Leu	Ser	Ala	Pro	Ser	Leu
					50					55					60
15	Gly	Ala	Val	Thr	Ala	Pro	Leu	Arg	Arg	Ala	Val	Asp	Tyr	Leu	Ala
					65					70					75
	Gly	Gly	Ala	Ala	Leu	Cys	Ser	Ala	Leu	Tyr	Val	Gly	Asp	Val	Cys
					80					85					90
	Gly	Ala	Leu	Phe	Leu	Val	Gly	Gln	Met	Phe	Thr	Tyr	Arg	Pro	Arg
					95					100					105
	Gln	His	Ala	Thr	Val	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Ser	Gly
					110					115					120
20	His	Ile	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp
					125					130					135
	Ser	Pro	Ala	Thr	Ala	Leu	Val	Met	Ala	Gln	Met	Leu	Arg	Ile	Pro
					140					145					150
	Gln	Val	Val	Ile	Asp	Ile	Ile	Ala	Gly	Gly	His	Trp	Gly	Val	Leu
					155					160					165
	Phe	Ala	Ala	Ala	Tyr	Phe	Ala	Ser	Ala	Ala	Asn	Trp	Ala	Lys	Val
					170					175					180
25	Val	Leu	Val	Leu	Phe	Leu	Phe	Ala	Gly	Val	Asp	Ala			
					185					190					

## (2) INFORMATION FOR SEQ ID NO:100:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA7

- 148 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

	Val	Pro	Tyr	Arg	Asn	Ala	Ser	Gly	Val	Tyr	His	Val	Thr	Asn	Asp	
					5					10					15	
	Cys	Pro	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Asp	Asn	Leu	Ile	Leu	
					20					25					30	
5	His	Ala	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Gln	Asn	Asn	Val	Ser	
					35					40					45	
	Arg	Cys	Trp	Val	Gln	Ile	Thr	Pro	Thr	Leu	Ser	Ala	Pro	Asn	Leu	
					50					55					60	
	Gly	Ala	Val	Thr	Ala	Pro	Leu	Arg	Arg	Ala	Val	Asp	Tyr	Leu	Ala	
					65					70					75	
	Gly	Gly	Ala	Ala	Leu	Cys	Ser	Ala	Leu	Tyr	Val	Gly	Asp	Ala	Cys	
					80					85					90	
10	Gly	Ala	Val	Phe	Leu	Val	Gly	Gln	Met	Phe	Ser	Tyr	Arg	Pro	Arg	
					95					100					105	
	Gln	His	Thr	Thr	Val	Gln	Asp	Cys	Asn	Cys	Ser	Ile	Tyr	Ser	Gly	
					110					115					120	
	His	Ile	Thr	Gly	His	Arg	Met	Ala	Trp	Asp	Met	Met	Met	Asn	Trp	
					125					130					135	
	Ser	Pro	Thr	Thr	Ala	Leu	Val	Met	Ala	Gln	Leu	Leu	Arg	Ile	Pro	
					140					145					150	
15	Gln	Val	Val	Ile	Asp	Ile	Ile	Ala	Gly	Gly	His	Trp	Gly	Val	Leu	
					155					160					165	
	Phe	Ala	Ala	Ala	Tyr	Phe	Ala	Ser	Ala	Ala	Asn	Trp	Ala	Lys	Val	
					170					175					180	
	Val	Leu	Val	Leu	Phe	Leu	Phe	Ala	Gly	Val	Asp	Ala				
					185					190						

20

## (2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 192 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

25

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA13

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

30	Val	Pro	Tyr	Arg	Asn	Ala	Ser	Gly	Val	Tyr	His	Val	Thr	Asn	Asp	
					5					10					15	
	Cys	Pro	Asn	Ser	Ser	Ile	Val	Tyr	Glu	Ala	Asp	Asp	Leu	Ile	Leu	
					20					25					30	
	His	Ala	Pro	Gly	Cys	Val	Pro	Cys	Val	Arg	Gln	Gly	Asn	Val	Ser	
					35					40					45	
	Arg	Cys	Trp	Val	Gln	Ile	Thr	Pro	Thr	Leu	Ser	Ala	Pro	Ser	Leu	
					50					55					60	

35

- 149 -

Gly Ala Val Thr Ala Pro Leu Arg Arg Ala Val Asp Tyr Leu Ala  
                                 65                                70                                75  
 Gly Gly Ala Ala Leu Cys Ser Ala Leu Tyr Val Gly Asp Ala Cys  
                                 80                                85                                90  
 Gly Ala Val Phe Leu Val Gly Gln Met Phe Thr Tyr Ser Pro Arg  
                                 95                                100                                105  
 Arg His Asn Val Val Gln Asp Cys Asn Cys Ser Ile Tyr Ser Gly  
 5                                110                                115                                120  
 His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                                 125                                130                                135  
 Ser Pro Thr Thr Ala Leu Val Met Ala Gln Leu Leu Arg Ile Pro  
                                 140                                145                                150  
 Gln Val Val Ile Asp Ile Ile Ala Gly Ala His Trp Gly Val Leu  
                                 155                                160                                165  
 10 Phe Ala Ala Ala Tyr Tyr Ala Ser Ala Ala Asn Trp Ala Lys Val  
                                 170                                175                                180  
 Val Leu Val Leu Phe Leu Phe Ala Gly Val Asp Ala  
                                 185                                190

## (2) INFORMATION FOR SEQ ID NO:102:

15 (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 192 amino acids  
       (B) TYPE: amino acid  
       (C) STRANDEDNESS: unknown  
       (D) TOPOLOGY: unknown

20 (vi) ORIGINAL SOURCE:  
       (A) ORGANISM: homosapiens  
       (C) INDIVIDUAL ISOLATE: HK2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:  
 Leu Thr Tyr Gln Asn Ser Ser Gln Leu Tyr His Leu Thr Asn Asp  
                                 1                                10                                15  
 Cys Pro Asn Ser Ser Ile Val Leu Glu Ala Asp Ala Met Ile Leu  
 25                                20                                25                                30  
 His Leu Pro Gln Cys Leu Pro Cys Val Arg Val Asp Asp Arg Ser  
                                 35                                40                                45  
 Thr Cys Trp His Ala Val Thr Pro Thr Leu Ala Ile Pro Asn Ala  
                                 50                                55                                60  
 Ser Thr Pro Ala Thr Gln Phe Arg Arg His Val Asp Leu Leu Ala  
                                 65                                70                                75  
 30 Gln Ala Ala Val Val Cys Ser Ser Leu Tyr Ile Gln Asp Leu Cys  
                                 80                                85                                90  
 Gln Ser Leu Phe Leu Ala Gln Gln Leu Phe Thr Phe Gln Pro Arg  
                                 95                                100                                105  
 Arg His Trp Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Thr Gln  
                                 110                                115                                120  
 His Val Thr Gln His Arg Met Ala Trp Asp Met Met Met Asn Trp  
                                 125                                130                                135  
 35 Ser Pro Thr Thr Thr Leu Val Leu Ser Ser Ile Leu Arg Val Pro  
                                 140                                145                                150

- 150 -

Glu Ile Cys Ala Ser Val Ile Phe Gln Gln His Trp Gln Ile Leu  
 155 160 165  
 Leu Ala Val Ala Tyr Phe Gln Met Ala Gln Asn Trp Leu Lys Val  
 170 175 180  
 Leu Ala Val Leu Phe Leu Phe Ala Gln Val Glu Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 103:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK7

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

15 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39  
 AAC ACC AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT 78  
 GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC 117  
 AGG GGC CCT AGA TTG GGT GTG CGC GCG CCG AGG AAG ACT 156  
 TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC 195  
 CCC AAG GCA CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG 234  
 20 CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC 273  
 GGG TGG GCG GGA TGG CTC CTG TCT CCC CGT GGC TCT CGG 312  
 CCT AGC TGG GGC CCC ACA GAC CCC CGG CGC AGG TCG CGC 351  
 AAT TTG GGT AAA GTC ATC GAT ACC CTT ACG TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATA CCG CTC GTC GGC GCC CCT 429  
 CTT GGA GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CCG 468  
 GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAC CTT 507  
 25 CCT GGT TGC TCT TTC TCT ATC TTC CTT TTG GCC CTG CTC 546  
 TCT TGC CTG ACC GTG CCC GCT TCG GCC 573

## (2) INFORMATION FOR SEQ ID NO: 104:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US11

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

- 151 -

5  
 10  
 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39  
 AAC ACC AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT 78  
 GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC 117  
 AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT 156  
 TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC 195  
 CCC AAG GCA CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG 234  
 CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC 273  
 GGG TGG GCG GGA TGG CTC CTG TCT CCC CGT GGC TCT CGG 312  
 CCT AGC TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGC 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTT ACG TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATA CCG CTC GTC GGC GCC CCT 429  
 CTC GGA GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG 468  
 GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAC CTT 507  
 CCT GGT TGC TCT TTC TCT ATC TTC CTT CTG GCC CTG CTC 546  
 TCT TGC CTG ACT GTG CCC GCT TCA GCC 573

## (2) INFORMATION FOR SEQ ID NO: 105:

15  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S14

20  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

25  
 30  
 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39  
 AAC ACC AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT 78  
 GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC 117  
 AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT 156  
 TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC 195  
 CCC AAG GCA CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG 234  
 CCC GGG TAT CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC 273  
 GGG TGG GCG GGA TGG CTC CTG TCT CCC CGT GGC TCT CGG 312  
 CCT AGC TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGC 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATA CCG CTC GTC GGC GCC CCC 429  
 CTC GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG 468  
 GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAC CTT 507  
 CCT GGT TGC TCT TTC TCT ATC TTC CTC CTA GCC CTG CTT 546  
 TCT TGC CTG ACT GTG CCC GCT TCA GCC 573

## (2) INFORMATION FOR SEQ ID NO: 106:

35  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid

- 152 -

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SW1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGT	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGT	78
	GGC	GGT	CAG	ATC	GTT	GGT	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCT	AGA	TTG	GGT	GTG	CGC	GCG	ACG	AGG	AAG	ACT	156
10	TCC	GAG	CGG	TCG	CAA	CCT	CGA	GGT	AGA	CGT	CAG	CCT	ATC	195
	CCC	AAG	GCG	CGT	CGG	CCC	GAG	GGC	AGG	ACC	TGG	GCT	CAG	234
	CCC	GGG	TAT	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TGC	273
	GGA	TGG	GCG	GGA	TGG	CTC	CTG	TCC	CCC	CGT	GGC	TCT	CGG	312
	CCT	AGC	TGG	GGC	CCT	ACA	GAC	CCC	CGG	CGT	AGG	TCG	CGC	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTC	ACG	TGC	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATT	CCG	CTC	GTC	GGC	GCC	CCT	429
	CTT	GGA	GGC	GCT	GCC	AGG	GCC	CTG	GCG	CAT	GGC	GTC	CGG	468
15	GTT	CTG	GAA	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAC	CTT	507
	CCT	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTT	CTG	GCC	CTG	CTT	546
	TCT	TGC	CTG	ACA	GTG	CCC	GCG	TCA	GCC					573

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: S18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

	ATG	AGC	ACA	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGT	CGC	CCA	CAG	GAC	GTT	AAG	TTC	CCG	GGT	78
	GGC	GGT	CAG	ATC	GTT	GGT	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCT	AGA	TTG	GGT	GTG	CGC	GCG	ACG	AGG	AAG	ACT	156
30	TCC	GAG	CGG	TCG	CAA	CCT	CGC	GGT	AGA	CGT	CAG	CCT	ATC	195
	CCC	AAG	GCG	CGT	CGG	CCC	GAG	GGC	AGG	ACC	TGG	GCT	CAG	234
	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TGC	273
	GGG	TGG	GCG	GGA	TGG	CTC	CTG	TCC	CCC	CGT	GGC	TCC	CGG	312
	CCT	AGC	TGG	GGC	CCT	ACA	GAC	CCC	CGG	CGT	AGG	TCG	CGC	351
	AAT	TTG	GGC	AAA	GTC	ATC	GAT	ACC	CTC	ACG	TGC	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATT	CCG	CTC	GTC	GGC	GCC	CCT	429
35	CTC	GGA	GGC	GCT	GCC	AGG	GCC	CTG	GCG	CAT	GGC	GTC	CGG	468
	GTT	CTG	GAA	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAC	CTT	507

- 153 -

CCT GGT TGC TCT TTC TCT ATC TTC CTT CTG GCC CTG CTC 546  
TCT TGT CTG ACT GTG CCC GCG TCA GCT 573

## (2) INFORMATION FOR SEQ ID NO: 108:

5

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 573 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10

- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: DR4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
	AAC ACC AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT	78
	GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC	117
15	AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT	156
	TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC	195
	CCC AAG GCG CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG	234
	CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC	273
	GGG TGG GCG GGA TGG CTC CTG TCC CCC CGT GGC TCT CGG	312
	CCT AGC TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGC	351
	AAT TTG GGT AAG GTC ATC GAC ACC CTC ACG TGC GGC TTC	390
20	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCC CCC	429
	CTT GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGA	468
	GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAT CTT	507
	CCT GGT TGC TCT TTC TCT ATC TTC CTT TTG GCT TTG CTC	546
	TCT TGC TTG ACC GTG CCC GCA TCG GCC	573

## (2) INFORMATION FOR SEQ ID NO: 109:

25

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 573 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30

- (vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: SA10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
35	GGT GGT CAG ATC GTT GGT GGA GTC TAT CTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACG AGG AAG ACT	156

- 154 -

TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC 195  
 CCC AAG GCT CGC CAG CCC GAG GGC AGG ACC TGG GCC CAG 234  
 CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TTG 273  
 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCT CGG 312  
 CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGT 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC 390  
 5 GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCT 429  
 TTA GGG GGC GCT GCC AGG GCC TTG GCG CAT GGC GTC CGG 468  
 GTT CTG GAA GAC GGC GTG AAC TAT GCA ACA GGG AAT TTG 507  
 CCC GGT TGC CCT TTC TCT ATC TTC CTC TTG GCT TTG CTG 546  
 TCC TGT TTA ACC ATC CCA GCT TCC GCT 573

10 (2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

20 ATG AGC ACG AAT CCT AAA CCT CAA AGA CAA ACC AAA CGT 39  
 AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGT 78  
 GGC GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC 117  
 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT 156  
 TCC GAG CGG TCA CAA CCT CGT GGA CGG CGA CAA CCT ATC 195  
 CCC AAG GCT CGC CGG CCC GAG GGC AGG GCC TGG GCC CAG 234  
 CCC GGG CAT CCT TGG CCC CTC TAT GGC AAT GAG GGC TTG 273  
 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCC CGG 312  
 CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGC 351  
 25 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC 429  
 CTA GGG GGC GCT GCC AGA GCC TTG GCG CAT GGC GTC CGG 468  
 GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT CTG 507  
 CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT CTG CTG 546  
 TCC TGC TTG ACC ATC CCA GCT TCC GCT 573

30 (2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (vi) ORIGINAL SOURCE:



- 155 -

(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: D1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
5	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT	156
	TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC	195
	CCC AAG GCT CGC CGG CCC GAG GGT AGG GCC TGG GCT CAG	234
	CCC GGG TAC CCT TGG CCC CTC TAT GGC AAC GAG GGC TTG	273
	GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCC CGG	312
10	CCT AGT TGG GGC CCC ACC GAC CCC CGG CGT AGG TCG CGT	351
	AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC	390
	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCC CCC	429
	CTA GGG GGT GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG	468
	GTT CTG GAG GAC GGC GTG AAT TAT GCA ACA GGG AAT TTG	507
	CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG CTG	546
	TCC TGT TTG ACC ATC CCA GCT TCC GCT	573

15

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 573 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: US6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

25	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT	156
	TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC	195
	CCC AAG GCT CGC CGG CCC GAG GGC AGG GCC TGG GCT CAG	234
	CCC GGG TAC CCT TGG CCC CTC TAT GGC AAC GAG GGC ATG	273
	GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCC CGG	312
30	CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGT	351
	AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC	390
	GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC	429
	CTA GGG GGC GCT GCC AGG GCC TTG GCG CAT GGC GTC CGG	468
	GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAC TTG	507
	CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG CTG	546
	TCC TGT TTG ACC ATT CCA GCT TCC GCT	573

35

- 156 -

## (2) INFORMATION FOR SEQ ID NO: 113:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: P10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

10	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT	156
	TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC	195
	CCC AAG GCT CGC CGG CCC GAG GGC AGG GCC TGG GCT CAG	234
	CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TTG	273
15	GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCT CGG	312
	CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGT	351
	AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC	390
	GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC	429
	CTA GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG	468
	GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT CTG	507
	CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG CTG	546
20	TCC TGC CTG ACC ATC CCA GCG TCC GCT	573

## (2) INFORMATION FOR SEQ ID NO: 114:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

30	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT	156
	TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC	195
	CCC AAG GCT CGC CGG CCC GAG GGC AGG GCC TGG GCT CAG	234
35	CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC ATG	273
	GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCT CGG	312

- 157 -

CCT AGT TGG GGC CCC AAC GAC CCC CGG CGT AGG TCG CGT 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC 429  
 CTA GGG GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG 468  
 GTT CTG GAG GAC GGC GTG AAC TAC GCA ACA GGG AAT TTG 507  
 CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT CTG TTG 546  
 5 TCC TGT TTG ACC ATC CCA GCT TCC GCC 573

## (2) INFORMATION FOR SEQ ID NO: 115:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 10  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T10  
 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39  
 AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC 78  
 GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC 117  
 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT 156  
 TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAG CCT ATC 195  
 CCC AAG GCT CGC CAG CCC GAG GGC AGG GCC TGG GCT CAG 234  
 20 CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC ATG 273  
 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGT GGC TCC CGG 312  
 CCT AGT TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGT 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC 429  
 CTA GGG GGC GCT GCC AGG GCT CTG GCA CAT GGT GTC CGG 468  
 GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT TTG 507  
 25 CCC GGT TGC TCT TTT TCT ATC TTC CTC TTG GCT CTG CTG 546  
 TCT TGT CTG ACC ATC CCA GCT TCC GCT 573

## (2) INFORMATION FOR SEQ ID NO: 116:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 30  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW2  
 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

- 158 -

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	78
	GGT	GGC	CAG	ATC	GTT	GGT	GGA	GTT	TAC	CTG	TTG	CCG	CGC	117
	AGG	GGC	CCC	CGG	TTG	GGT	GTG	CGC	CCG	ACT	AGG	AAG	ACT	156
	TCC	GAG	CGG	TCG	CAA	CCT	CGT	GGA	AGG	CGA	CAA	CCT	ATC	195
5	CCC	AAG	GCT	CGC	CAG	CCC	GAG	GGC	AGG	GCC	TGG	GCT	CAG	234
	CCT	GGG	TAC	CCC	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	ATG	273
	GGA	TGG	GCA	GGA	TGG	CTC	CTG	TCC	CCC	CGC	GGC	TCT	CGG	312
	CCT	AGT	TGG	GGC	CCC	ACT	GAC	CCC	CGG	CGT	AGG	TCG	CGT	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTC	ACA	TGC	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATT	CCG	CTC	GTC	GGC	GCC	CCC	429
	CTA	GGG	GGC	GCT	GCC	AGG	GCC	CTG	GCG	CAT	GGC	GTC	CGG	468
10	GTC	CTG	GAG	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAT	CTG	507
	CCC	GGT	TGC	TCC	TTT	TCT	ATC	TTC	CTC	TTG	GCT	TTG	CTG	546
	TCC	TGT	CTG	ACC	ATC	CCA	GCT	TCC	GCT					573

## (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: IND3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	78
	GGT	GGC	CAG	ATC	GTT	GGT	GGA	GTT	TAC	CTG	TTG	CCG	CGC	117
	AGG	GGC	CCC	AGG	TTG	GGT	GTG	CGC	GCG	ACT	AGG	AAG	ACT	156
	TCC	GAG	CGG	TCG	CAA	CCT	CGT	GGA	AGG	CGA	CAA	CCT	ATC	195
25	CCC	AAG	GCT	CGC	CGG	CCC	GAG	GGT	AGG	GCC	TGG	GCT	CAG	234
	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TTG	273
	GGG	TGG	GCA	GGA	TGG	CTC	CTG	TCA	CCC	CGC	GGT	TCT	CGG	312
	CCT	AGT	TGG	GGC	CCC	ACA	GAC	CCC	CGG	CGT	AGG	TCG	CGT	351
	AAT	TTG	GGT	AAA	GTC	ATC	GAT	ACC	CTC	ACA	TGC	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATC	CCG	CTC	GTC	GGC	GCC	CCC	429
	CTA	GGG	GGC	GCT	GCC	AGG	GCC	CTG	GCG	CAT	GGC	GTC	CGG	468
30	GTC	CTG	GAG	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAC	TTG	507
	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTT	TTA	GCT	TTG	CTA	546
	TCC	TGT	TTG	ACC	ATC	CCA	GCT	TCC	GCT					573

## (2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs

- 159 -

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: IND8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	78
	GGT	GGC	CAG	ATC	GTT	GGT	GGA	GTT	TAC	CTG	TTG	CCG	CGC	117
10	AGG	GGC	CCC	AGG	TTG	GGT	GTG	CGC	GCG	ACT	AGG	AAG	ACT	156
	TCC	GAG	CGG	TCG	CAA	CCT	CGT	GGA	AGG	CGA	CAA	CCT	ATC	195
	CCC	AAG	GCT	CGC	CGG	CCC	GAG	GGT	AGG	GCC	TGG	GCT	CAG	234
	CCC	GGG	CAC	CCT	TGG	CCC	CTC	TAT	GGC	AAT	GAG	GGC	TTG	273
	GGG	TGG	GCA	GGA	TGG	CTC	CTG	TCA	CCC	CGC	GGC	TCT	CGG	312
	CCT	AGT	TGG	GGC	CCC	ACA	GAC	CCC	CGG	CGT	AGG	TCG	CGT	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTC	ACA	TGC	GGC	TTT	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATC	CCG	CTC	GTC	GGC	GCC	CCC	429
15	CTA	GGG	GGT	GCT	GCC	AGG	GCC	CTG	GCG	CAT	GGC	GTC	CGG	468
	GTC	CTG	GAG	GAC	GGC	GTG	AAC	TAT	GCA	ACA	GGG	AAC	TTG	507
	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTT	TTG	GCT	TTG	CTA	546
	TCC	TGT	TTG	ACC	GTC	CCA	GCT	TCC	GCT					573

## (2) INFORMATION FOR SEQ ID NO: 119:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S9

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTT	AAG	TTC	CCG	GGC	78
	GGT	GGT	CAG	ATC	GTC	GGT	GGA	GTT	TAC	CTG	TTG	CCG	CGC	117
30	AGG	GGC	CCC	AGG	TTG	GGT	GTG	CGC	GCA	ACT	AGG	AAG	ACT	156
	TCC	GAG	CGG	TCG	CAA	CCT	CGT	GGA	AGG	CGA	CAA	CCT	ATC	195
	CCC	AAG	GCT	CGC	CAT	CCC	GAG	GGC	AGG	GCC	TGG	GCT	CAG	234
	CCC	GGG	TAC	CCT	TGG	CCC	CTC	TAC	GGC	AAT	GAG	GGC	TTG	273
	GGG	TGG	GCA	GGA	TGG	CTC	CTG	TCA	CCC	CGT	GGC	TCT	CGG	312
	CCT	AGT	TGG	GGC	CCC	AAT	GAC	CCC	CGG	CGT	AGG	TCG	CGT	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTC	ACA	TGC	GGC	TTT	390
35	GCC	GAC	CTC	ATG	GGG	TAC	ATT	CCG	CTC	GTC	GGC	GCC	CCC	429
	CTA	GGG	GGC	GCT	GCC	AGG	GCT	CTG	GCG	CAT	GGC	GTC	CGG	468

- 160 -

GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAC CTC	507
CCC GGT TGC TCT TTC TCT ATC TTC CTT CTG GCT TTG CTG	546
TCC TGT TTG ACC ATC CCA GCT TCC GCT	573

5 (2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
15 GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117
AGG GGC CCC AGG TTG GGT GTG CGC GCG ACC AGG AAG ACT	156
TCA GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC	195
CCC AAG GCT CGC CAA CCC GAG GGC AGG ACC TGG GCT CAG	234
CCC GGG TAT CCT TGG CCC CTC TAT GGC AAC GAG GGC ATG	273
GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCT CGG	312
CCT AAT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGC	351
20 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC	390
GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGT GCC CCC	429
CTA GGG GGC GTT GCC AGA GCC TTG GCA CAT GGT GTC CGG	468
GTT CTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT TTA	507
CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG CTG	546
TCC TGC TTG ACC ACC CCA GCT TCC GCT	573

25 (2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
35 AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117

- 161 -

AGG GGC CCC AGG TTG GGT GTG CGC GCG ACC AGG AAG ACT 156  
 TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC 195  
 CCC AAG GCT CGC CGA CCC GAG GGC AGG ACC TGG GCT CAG 234  
 CCC GGG TAT CCT TGG CCC CTC TAT GGC AAT GAG GGC ATG 273  
 GGG TGG GCA GGA TGG CTC CTG TCA CCC CAT GGC TCT CGG 312  
 CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGT 351  
 5 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCC CCC 429  
 CTA GGG GGC GTT GCC AGA GCC CTG GCA CAC GGT GTC CGG 468  
 GTT CTG GAG GAC GGC GTG AAC TAC GCA ACA GGG AAT ATA 507  
 CCC GGT TGC TCT TTC TCT ATC TTC CTT TTG GCT TTG CTG 546  
 TCC TGT CTG ACC ACC CCA GTT TCC GCT 573

10 (2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

20 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAG ACC AAA CGT 39  
 AAC ACC AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGC 78  
 GGT GGC CAG ATC GTC GGT GGA GTT TAC CTG TTG CCG CGC 117  
 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT 156  
 TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC 195  
 CCC AAG GCT CGC CAA CCC GAG GGC AGG ACC TGG GCT CAG 234  
 CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC ATG 273  
 GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCT CGG 312  
 25 CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGC 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC 390  
 GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCC CCC 429  
 TTA GGG GGC GTT GCC AGA GCC CTG GCA CAT GGT GTC CGG 468  
 GTT GTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT TTG 507  
 CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT CTG CTG 546  
 TCC TGT TTG ACC ATC CCA GCT TCC GCT 573

30

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35

- 162 -

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: P8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

5	ATG AGC ACG ACT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
	AAC ACC AGC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT	156
	TCC GAG CGA TCG CAA CCT CGT GGC AGG CGA CAA CCT ATC	195
	CCC AAG GCT CGC CGG CCC GAG GGT AGG GCC TGG GCT CAG	234
	CCC GGG CAC CCT TGG CCC CTC TAT GCC AAT GAG GGC TTG	273
10	GGG TGG GCG GGA TGG CTC CTG TCA CCC CGC GGC TCC CGG	312
	CCT AGT TGG GGC CCC ACG GAC CCC CGG CGT AGG TCG CGC	351
	AAT TTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC	390
	GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GGC CCC	429
	CTA GGG GGC GTT GCC AGG GCC CTG GCG CAT GGC GTC CGG	468
	GTT GTG GAG GAC GGC GTG AAC TAT GCA ACA GGG AAT CTG	507
	CCT GGT TGC TCT TTC TCT ATC TTC CTT TTG GCT TTG CTG	546
15	TCT TGT CTG ACC ATC CCA GCT TCC GCT	573

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

25	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC CTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT AGG AAG ACT	156
	TCC GAG CGG TCG CAA CCT CGT GGA AGG CGA CAA CCT ATC	195
	CCC AAG GCT CGC CGG CCC GAG GGT AGG GCC TGG GCT CAG	234
	CCC GGG TAC CCT TGG CCC CTC TAT GGC GAG GGC ATG	273
30	GGG TGG GCA GGA TGG CTC CTG TCA CCC CGC GGC TCC CGG	312
	CCT AAT TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG CGT	351
	AAT CTG GGT AAG GTC ATC GAT ACC CTC ACA TGC GGC TTC	390
	GCC GAC CTC ATG GGG TAC ATT CCG CTC GTC GGC GCT CCC	429
	TTA GGG GGC GTT GCC AGG GCC CTG GCG CAT GGC GTC CGG	468
	GTT CTG GAG GAC GGC GTG AAT TAC GCA ACA GGG AAT TTG	507
	CCT GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG CTG	546
35	TCC TGC TTG ACC ATC CCA GCT TCC GCT	573



- 163 -

## (2) INFORMATION FOR SEQ ID NO: 125:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T4

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

10	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC AAC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGC	78
	GGC GGC CAG ATC GTT GGC GGA GTA TAC TTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACA AGG AAG ACT	156
	TCG GAG CGA TCC CAG CCA CGT GGG AGG CGC CAG CCC ATC	195
	CCC AAA GAT CGG CGC TCC ACT GGC AAG TCC TGG GGA AAA	234
	CCA GGA TAT CCC TGG CCC CTG TAT GGG AAT GAG GGA CTC	273
15	GGC TGG GCA GGA TGG CTC CTG TCC CCC CGA GGT TCC CGT	312
	CCC TCC TGG GGC CCC AAT GAC CCC CGG CAT AGG TCG CGC	351
	AAC GTG GGT AAG GTC ATC GAT ACC CTA ACG TGC AGC CTT	390
	GCC GAC CTC ATG GGG TAC GTC CCC CTC GTA GGC GGC CCG	429
	TTG GGT GGC GTC GCC AGA GCT CTC GCG CAT GGC GTG AGA	468
	GTC CTG GAG GAC GGG GTT AAT TAT GCA ACA GGG AAC TTA	507
	CCT GGT TGC TCC TTT TCT ATT TTC TTG CTG GCC CTA CTG	546
20	TCC TGC ATC ACC ATT CCA GTC TCC GCT	573

## (2) INFORMATION FOR SEQ ID NO: 126:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

30	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACT AAC CGT CGC CCA CAA GAC GTT AAG TTT CCG GGC	78
	GGC GGC CAG ATC GTT GGC GGA GTA TAC TTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACA AGG AAG ACT	156
	TCG GAG CGG TCC CAG CCA CGT GGG AGG CGC CAG CCC ATC	195
35	CCC AAA GAT CGG CGC CCC ACT GGC AAG TCC TGG GGA AAA	234
	CCA GGA TAC CCT TGG CCC CTA TAT GGG AAT GAG GGA CTC	273

- 164 -

GGC TGG GCA GGA TGG CTC CTG TCC CCC CGA GGT TCC CGT 312  
 CCC TCT TGG GGC CCC ACT GAT CCC CGG CAT AGG TCG CGC 351  
 AAC GTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGC TTT 390  
 GCC GAC CTC ATG GGA TAC ATC CCC GTC GTG GGC GCT CCG 429  
 CTT GGT GGC GTC GCC AGA GCT CTC GCG CAT GGC GTG AGG 468  
 GTC CTG GAG GAC GGG GTT AAT TAT GCA ACA GGG AAC TTA 507  
 CCC GGT TGC TCC TTT TCT ATC TTC TTG CTG GCC TTA CTG 546  
 TCC TGC ATC ACC ATT CCA GTC TCT GCT 573

## (2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

ATG AGC ACA AAT CCA AAA CCC CAA AGA AAA ACC ATA AGA 39  
 AAC ACC AAC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGC 78  
 GGC GGC CAG ATC GTT GGC GGA GTA TAC TTG TTG CCG CGC 117  
 AGG GGC CCT AGG TTG GGT GTG CGC ACG ACA AGG AAG ACT 156  
 TCG GAG CGG TCC CAG CCA CGT GGG AGG CGC CAG CCC ATC 195  
 CCC AAA GAT CGG CGC TCC ACT GGC AAG TCC TGG GGA AAA 234  
 CCA GGA TAC CCC TGG CCT CTA TAT GGG AAT GAG GGA CTC 273  
 GGC TGG GCG GGA TGG CTC CTG TCC CCC CGA GGT TCC CGT 312  
 CCC TCT TGG GGC CCC AGT GAC CCC CGG CAT AGG TCG CGC 351  
 AAC GTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGC TTT 390  
 GCC GAC CTC ATG GGG TAC ATC CCC GTC GTA GGC GCC CCG 429  
 CTT GGT GGC GTT GCC AGA GCT CTC GCG CAC GGC GTG AGA 468  
 GTC CTG GAG GAC GGG GTT AAT TAT GCA ACA GGG AAC CTA 507  
 CCT GGT TGC TCT TTT TCT ATC TTC TTG CTG GCC CTA CTG 546  
 TCC TGC ATC ACC ACT CCG GCC TCT GCT 573

## (2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T2

- 165 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

	ATG AGC ACA ATT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACT AAC CGT CGC CCA CAA GAC GTT AAG TTT CCG GGC	78
	GGC GGC CAG ATC GTT GGC GGA GTA TAC TTG CTG CCG CGC	117
5	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACA AGG AAG ACT	156
	TCG GAG CGG TCC CAG CCT CGT GGA AGG CGC CAG CCC ATC	195
	CCT AAA GAT CGG CGC TCC ACT GGC AAG TCC TGG GGA AAA	234
	CCA GGA TAC CCC TGG CCC CTG TAT GGG AAT GAG GGG CTC	273
	GGC TGG GCA GGA TGG CTC CTG TCC CCC CGA GGT TCT CGT	312
	CCC TCT TGG GGC CCC AAT GAC CCC CGG CAT AGG TCG CGC	351
	AAT GTG GGT AAA GTC ATC GAT ACC CTA ACG TGC GGC TTT	390
	GCC GAC CTC ATG GGG TAC ATC CCC GTC GTA GGC GCC CCG	429
10	CTT GGT GGT GTC GCC AGA GCT CTT GCG CAT GGC GTG AGA	468
	GTC CTG GAG GAC GGA GTT AAT TAT GCA ACA GGT AAC TTA	507
	CCC GGT TGC TCC TTT TCT ATC TTC TTG CTA GCC CTG CTG	546
	TCC TGC ATC ACT ATT CCG GTT TCA GCT	573

## (2) INFORMATION FOR SEQ ID NO: 129:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACA AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGT	78
	GGC GGC CAG ATC GTT GGC GGA GTT TAC TTG CTG CCG CGC	117
25	AGG GGC CCT AGG TTG GGT GTG CGC GCG ACA AGG AAG ACT	156
	TCC GAG CGA TCC CAG CCG CGT GGG AGA CGC CAG CCC ATC	195
	CCG AAA GAT CGG CGC TCC ACC GGC AAG TCC TGG GGA AAA	234
	CCA GGA TAT CCT TGG CCT CTT TAC GGA AAC GAG GGC TGC	273
	GGT TGG GCA GGT TGG CTC CTG TCC CCC CGC GGG TCT CGT	312
	CCT ACT TGG GGC CCC ACT GAC CCC CGG CAT AGA TCA CGT	351
	AAT TTG GGC AGA GTC ATC GAT ACC ATT ACA TGT GGT TTT	390
	GCC GAC CTC ATG GGG TAC ATC CCT GTC GTT GGC GCC CCG	429
30	GTC GGA GGC GTC GCC AGA GCT CTG GCA CAT GGT GTT AGG	468
	GTC CTG GAA GAC GGG ATA AAC TAT GCA ACA GGG AAT TTG	507
	CCT GGT TGC TCT TTT TCT ATC TTC TTG CTT GCT CTT CTG	546
	TCA TGC TTC ACA GTG CCA GTG TCT GCA	573

## (2) INFORMATION FOR SEQ ID NO: 130:

35 (i) SEQUENCE CHARACTERISTICS:

- 166 -

(A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACA AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGT	78
10	GGC GGT CAG ATC GTT GGC GGA GTT TAC TTG CTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC ACA AGG AAG ACT	156
	TCC GAG CGA TCC CAG CCG CGT GGG AGA CGC CAG CCC ATC	195
	CCG AAA GAT CGG CGC TCC ACC GGC AAG TCC TGG GGA AAG	234
	CCA GGA TAT CCT TGG CCT CTG TAC GGA AAC GAG GGC TGC	273
	GGC TGG GCA GGT TGG CTC CTG TCC CCC CGC GGG TCT CGT	312
	CCT ACT TGG GGC CCC ACT GAC CCC CGG CAC AGA TCA CGT	351
	AAC TTG GGC AAG GTC ATC GAT ACC ATT ACG TGT GGT TTT	390
15	GCC GAC CTC ATG GGG TAC ATC CCT GTC GTT GGC GCC CCG	429
	GTC GGA GGC GTC GCC AGA GCT CTG GCA CAC GGT GTT AGG	468
	GTC CTG GAA GAC GGG ATA AAT TAC GCA ACA GGG AAT CTG	507
	CCT GGT TGC TCC TTT TCT ATC TTC TTA CTT GCT CTT CTG	546
	TCG TGC GCC ACG GTG CCG GTG TCT GCA	573

20 (2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

	ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAT ACA AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGT	78
30	GGC GGC CAG ATC GTT GGC GGA GTT TAC TTG CTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC ACG ACA AGG AAG ACT	156
	TCC GAG CGA TCC CAG CCG CGT GGG AGA CGC CAG CCC ATC	195
	CCG AAA GAT CGG CGC TCC ACC GGC AAG CCC TGG GGA AAG	234
	CCA GGA TAT CCT TGG CCC CTG TAT GGA AAC GAG GGC TGC	273
	GGC TGG GCA GGT TGG CTC CTG TCC CCC CGC GGG TCT CAT	312
	CCT AAT TGG GGC CCC ACT GAC CCC CGG CAT AAA TCA CGC	351
35	AAT TTG GGT AAA GTC ATC GAC ACC ATT ACG TGT GGT TTT	390
	GCC GAC CTC ATG GGG TAC ATC CCT GTC GTC GGC GCC CCG	429

- 167 -

GTC	GGA	GGC	GTC	GCC	AGA	GCT	CTG	GCA	CAC	GGT	GTT	AGA	468
GTC	CTG	GAA	GAC	GGG	ATA	AAT	TAC	GCA	ACA	GGG	AAT	CTG	507
CCT	GGT	TGC	TCT	TTT	TCT	ATC	TTC	TTA	CTT	GCT	CTT	CTG	546
TCA	TGC	TGC	ACA	GTG	CCA	GTG	TCT	GCG					573

## 5 (2) INFORMATION FOR SEQ ID NO: 132:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

ATG	AGC	ACA	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	AGA	39
AAT	ACA	AAC	CGC	CGC	CCA	CAG	GAC	GTT	AAG	TTC	CCG	GGT	78
GGC	GGC	CAG	ATC	GTT	GGC	GGA	GTT	TAC	TTG	CTG	CCG	CGC	117
AGG	GGC	CCC	AGG	TTG	GGT	GTG	CGC	GCG	ACA	AGG	AAG	ACT	156
TCC	GAG	CGA	TCC	CAG	CCG	CGT	GGG	AGA	CGC	CAG	CCC	ATC	195
CCG	AAA	GAT	CGG	CGC	TCC	ACC	GGC	AAG	TCC	TGG	GGA	AAG	234
CCA	GGA	TAT	CCT	TGG	CCC	CTG	TAT	GGA	AAC	GAG	GGC	TGC	273
GGC	TGG	GCA	GGT	TGG	CTC	CTG	TCC	CCC	CGC	GGG	TCT	CAT	312
CCT	AAT	TGG	GGC	CCC	ACT	GAC	CCC	CGG	CAT	AGA	TCA	CGC	351
AAT	TTG	GGC	AAA	GTC	ATC	GAC	ACC	ATT	ACG	TGT	GGT	TTT	390
GCC	GAC	CTC	ATG	GGG	TAC	ATC	CCT	GTC	GTT	GGC	GCC	CCG	429
GTC	GGA	GGC	GTC	GCC	AGA	GCT	CTG	GCA	CAC	GGT	GTT	AGA	468
GTC	CTG	GAA	GAC	GGG	ATA	AAT	TAC	GCA	ACA	GGG	AAT	CTG	507
CCT	GGT	TGC	TCT	TTT	TCT	ATC	TTC	TTA	CTT	GCT	CTT	CTG	546
TGC	TGC	TTC	ACA	GTG	CCA	GTG	TCT	GCG					573

## 25 (2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

ATG	AGC	ACA	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	AGA	39
AAC	ACA	AAC	CGC	CGC	CCA	CAG	GAC	GTT	AAG	TTC	CCG	GGT	78

- 168 -

GGC GGC CAG ATC GTT GGC GGA GTT TAC TTG CTG CCG CGC 117  
 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACA AGG AAG TCT 156  
 TCC GAG CGA TCC CAG CCG CGT GGG AGG CGC CAG CCC ATC 195  
 CCG AAA GAT CGG CGC TCC ACC GGC AAG TCC TGG GGA AAA 234  
 CCG GGA TAT CCT TGG CCC CTG TAT GGA AAC GAG GGC TGC 273  
 GGC TGG GCA GGT TGG CTC CTG TCC CCC CGC GGG TCT CGT 312  
 5 CCT ACT TGG GGC CCC ACT GAC CCC CGG CAT AGA TCA CGC 351  
 AAT TTG GGC AAA GTC ATC GAC ACC ATT ACG TGT GGT TTT 390  
 GCC GAC CTC ATG GGG TAC ATC CCT GTC GTT GGC GCC CCG 429  
 GTT GGA GGC GTC GCC AGA GCT CTG GCA CAC GGT GTT AGG 468  
 GTC CTG GAA GAC GGG ATA AAT TAC GCA ACA GGG AAT TTG 507  
 CCT GGT TGC TCT TTT TCT ATC TTC TTG CTT GCT CTT CTG 546  
 TCG TGC TGC ACA GTG CCA GTG TCT GCG 573

10

## (2) INFORMATION FOR SEQ ID NO: 134:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S83

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

20 ATG AGC ACA AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA 39  
 AAC ACT AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC 78  
 GGT GGC CAG ATC GTT GGC GGA GTA TAC TTG CTG CCG CGC 117  
 AGG GGC CCG AGA TTG GGT GTG CGC GCG ACG AGG AAA ACT 156  
 TCC GAA CGG TCC CAG CCA CGT GGG AGG CGC CAG CCC ATC 195  
 CCT AAA GAT CGG CGC ACC ACT GGC AAG TCC TGG GGA AGG 234  
 CCA GGA TAC CCT TGG CCC CTG TAT GGG AAT GAG GGC CTC 273  
 25 GGC TGG GCA GGG TGG CTC CTG TCC CCC CGC GGT TCT CGC 312  
 CCT TCA TGG GGC CCC ACC GAC CCC CGG CAT AAA TCG CGC 351  
 AAC TTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGT TTT 390  
 GCC GAC CTC ATG GGG TAC ATA CCC GTC GTT GGC GCT CCC 429  
 GTT GGC GGC GTT GCC AGA GCC CTC GCC CAT GGG GTG AGG 468  
 GTT CTG GAG GAC GGG ATA AAT TAT GCA ACG GGG AAT TTG 507  
 CCC GGT TGC TCT TTC TCT ATC TTT CTC TTG GCC CTC TTG 546  
 30 TCT TGC ATC TCT GTG CCA GTT TCC GCC 573

30

## (2) INFORMATION FOR SEQ ID NO: 135:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: singl  
 (D) TOPOLOGY: linear

35

- 169 -

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

5	ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGT	78
	GGC GGA CAG ATC GTT GGT GGA GTA TAC GTG TTG CCG CGC	117
	AGG GGC CCA CGA TTG GGT GTG CGC GCG ACG CGT AAA ACT	156
	TCT GAA CGG TCG CAG CCT CGC GGA CGA CGA CAG CCT ATC	195
	CCC AAG GCG CGT CGG AGC GAA GGC CGG TCC TGG GCT CAG	234
	CCC GGG TAC CCT TGG CCC CTC TAT GGT AAC GAG GGC TGC	273
10	GGG TGG GCA GGA TGG CTC CTG TCC CCA CGC GGC TCC CGT	312
	CCA TCT TGG GGC CCA AAC GAC CCC CGG CGA CGG TCC CGC	351
	AAT TTG GGT AAA GTC ATC GAT ACC CTT ACG TGC GGA TTC	390
	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCT CCC	429
	GTA GGA GGC GTC GCA AGA GCC CTC GCG CAT GGC GTG AGG	468
	GCC CTT GAA GAC GGG ATA AAT TTC GCA ACA GGG AAC TTG	507
	CCC GGT TGC TCC TTT TCT ATC TTC CTT CTT GCT CTG TTC	546
15	TCT TGC TTA ATT CAT CCA GCA GCT AGT	573

(2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S52

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

25	ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC ATC CGT CGC CCA CAG GAC GTT AAG TTC CCG GGT	78
	GGC GGA CAG ATC GTT GGT GGA GTA TAC GTG TTG CCG CGC	117
	AGG GGC CCA CGA TTG GGT GTG CGC GCG ACG CGT AAA ACT	156
	TCT GAA CGG TCA CAG CCT CGC GGA CGA CGA CAG CCT ATC	195
	CCC AAG GCG CGT CGG AGC GAA GGC CGG TCC TGG GCT CAG	234
	CCC GGG TAC CCT TGG CCC CTC TAT GGT AAT GAG GGC TGC	273
30	GGG TGG GCA GGG TGG CTC CTG TCC CCA CGC GGC TCC CGT	312
	CCA TCT TGG GGC CCA AAC GAC CCC CGG CGG AGG TCC CGC	351
	AAT TTG GGT AAA GTC ATC GAT ACC CTT ACG TGC GGA TTC	390
	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCT CCC	429
	GTA GGA GGC GTC GCA AGA GCC CTC GCG CAT GGC GTG AGG	468
	GCC CTT GAA GAC GGG ATA AAT TTT GCA ACA GGG AAC TTG	507
	CCC GGT TGC TCC TTT TCT ATC TTC CTT CTT GCT CTG TTC	546
35	TCC TGC TTA GTT CAT CCT GCA GCT AGT	573

- 170 -

## (2) INFORMATION FOR SEQ ID NO: 137:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

	ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC ATC CGT CGC CCA CAG GAC ATC AAG TTC CCG GGT	78
	GGC GGA CAG ATC GTT GGT GGA GTA TAC GTG TTG CCG CGC	117
	AGG GGC CCA CGA TTG GGT GTG CGC GCG ACG CGT AAA ACT	156
	TCT GAA CGG TCA CAG CCT CGC GGA CGG CGA CAG CCT ATC	195
	CCC AAG GCG CGT CGG AGC GAA GGC CGA TCC TGG GCT CAG	234
15	CCC GGG TAC CCT TGG CCC CTC TAT GGT AAC GAG GGC TGC	273
	GGG TGG GCA GGG TGG CTC CTG TCC CCA CGC GGC TCC CGT	312
	CCA TCT TGG GGC CCA AAT GAC CCC CGG CGG AGG TCC CGC	351
	AAT TTG GGT AAA GTC ATC GAT ACC CTT ACG TGC GGC TTC	390
	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCT CCC	429
	GTA GGA GGC GTC GCA AGA GCC CTC GCG CAT GGC GTG AGG	468
	GCC CTT GAA GAC GGG ATA AAT TTT GCA ACA GGG AAC TTG	507
20	CCC GGT TGC TCT TTT TCT ATC TTC CTT CTT GCC CTG TTC	546
	TCT TGC TTA ATT CAT CCA GCA GCT AGT	573

## (2) INFORMATION FOR SEQ ID NO: 138:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK12

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

	ATG AGC ACA CTT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC ATC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT	78
	GGC GGA CAG ATC GTT GGT GGA GTA TAC GTG TTG CCG CGC	117
	AGG GGC CCA CGA TTG GGT GTG CGC GCG ACG CGT AAA ACT	156
	TCT GAA CGG TCA CAG CCT CGC GGA CGG CGA CAG CCT ATC	195
35	CCC AAG GCG CGT CGG AGC GAA GGC CGG TCC TGG GCT CAG	234
	CCT GGG TAC CCT TGG CCC CTC TAT GGT AAC GAG GGC TGC	273



- 171 -

GGG TGG GCA GGG TGG CTC CTG TCC CCA CGC GGC TCC CGT 312  
 CCA TCT TGG GGC CCA AAC GAC CCC CGG CGG AGG TCC CGC 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTC ACG TGC GGA TTC 390  
 GCC GAC CTC ATG GGG TAC ATC CCG CTC GTC GGC GCT CCT 429  
 GTA GGG GGC GTC GCA AGA GCC CTC GCG CAT GGC GTG AGG 468  
 GCC CTT GAA GAC GGG ATA AAT TTC GCA ACA GGG AAC TTG 507  
 5 CCC GGT TGC TCC TTT TCT ATC TTC CTT CTT GCT CTG TTC 546  
 TCT TGC CTA ATT CAT CCA GCA GCT AGT 573

## (2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 10  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z4  
 15  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:  
 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT 39  
 AAC ACC AAC CGC CGC CCC ATG GAC GTA AAG TTC CCG GGT 78  
 GGT GGC CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC 117  
 AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT CGA AAG ACT 156  
 TCG GAG CGG TCG CAA CCT CGT GGC AGG CGT CAA CCT ATC 195  
 20 CCC AAG GCG CGC CAG CCA GAG GGC AGA TCC TGG GCG CAG 234  
 CCC GGG TAC CCT TGG CCC CTC TAT GGC AAT GAG GGC TGC 273  
 GGG TGG GCA GGG TGG CTC CTG TCT CCT CGC GGC TCT CGG 312  
 CCA TCT TGG GGC CCA AAT GAT CCC CGG CGG AGA TCG CGC 351  
 AAT CTG GGT AAG GTC ATC GAT ACC CTG ACG TGC GGC TTC 390  
 GCC GAC CTC ATG GGA TAC ATC CCG ATC GTG GGC GCC CCC 429  
 GTG GGG GGC GTC GCC AGG GCT CTG GCG CAT GGC GTC AGG 468  
 25 GCT GTG GAG GAC GGG ATT AAC TAT GCA ACA GGG AAT CTT 507  
 CCC GGT TGC TCT TTC TCT ATC TTC CTT TTG GCA CTT CTT 546  
 TCG TGC CTC ACT GTT CCA GCG TCG GCT 573

## (2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 30  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z8  
 35

- 172 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGC	CGC	CCT	ATG	GAT	GTA	AAA	TTC	CCA	GGC	78
	GGC	GGC	CAG	ATC	GTT	GGC	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCC	AGG	TTG	GGT	GTG	CGC	GCG	ACT	CGG	AAG	ACT	156
5	TCG	GAG	CGG	TCG	CAA	CCT	CGT	GGC	AGG	CGT	CAG	CCT	ATC	195
	CCC	AAG	GCA	CGT	CGG	TCC	GAG	GGT	AGG	TCC	TGG	GCT	CAG	234
	CCC	GGG	TAC	CCA	TGG	CCT	CTT	TAC	GGT	AAT	GAA	GGC	TGT	273
	GGG	TGG	GCA	GGT	TGG	CTC	CTG	TCC	CCC	CGC	GGC	TCT	CGA	312
	CCG	TCT	TGG	GGC	CCA	AAT	GAT	CCC	CGG	CGG	AGG	TCG	CGC	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTC	ACG	TGC	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGA	TAC	ATC	CCG	CTC	GTG	GGC	GCC	CCA	429
10	GTA	GGA	GGC	GTC	GCC	AGA	GCC	CTG	GCG	CAT	GGC	GTC	AGG	468
	GCT	GTG	GAG	GAC	GGG	ATC	AAC	TAT	GCA	ACA	GGG	AAC	CTT	507
	CCT	GGT	TGC	TCT	TTC	TCT	ATC	TTC	CTC	TTG	GCA	CTT	CTC	546
	TCG	TGC	CTA	ACC	GTC	CCA	GCG	TCT	GCT					573

## (2) INFORMATION FOR SEQ ID NO: 141:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

	ATG	AGC	ACA	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGT	CGC	CCC	ATG	GAT	GTG	AAA	TTC	CCG	GGC	78
	GGC	GGC	CAG	ATC	GTT	GGC	GGA	GTT	TAC	TTG	CTG	CCG	CGC	117
25	AGG	GGC	CCC	CGG	TTG	GGT	GTG	CGC	GCA	GCT	CGG	AAG	ACT	156
	TCG	GAG	CGG	TCA	CAA	CCT	CGT	GGC	AGG	CGT	CAG	CCT	ATC	195
	CCC	AAG	GCG	CGC	CGG	TCC	GAG	GGC	AGG	TCC	TGG	GCT	CAG	234
	CCC	GGG	TAC	CCT	TGG	CCC	CTT	TAC	GGC	AAT	GAG	GGC	TGT	273
	GGG	TGG	GCA	GGG	TGG	CTC	CTG	TCC	CCC	CGC	GGT	TCC	AGG	312
	CCG	TCT	TGG	GGC	CCC	AAT	GAT	CCC	CGG	CGT	AGG	TCC	CGT	351
	AAT	CTG	GGT	AAA	GTC	ATC	GAT	ACC	CTG	ACG	TGT	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGA	TAC	ATT	CCG	CTC	GTA	GGC	GCC	CCT	429
30	GTG	GGT	GGC	GTC	GCC	AGG	GCC	CTG	GCG	CAT	GGC	GTC	AGG	468
	GCC	GTG	GAG	GAC	GGA	ATT	AAC	TAC	GCA	ACA	GGG	AAC	CTT	507
	CCT	GGT	TGC	TCT	TTC	TCT	ATC	TTT	CTT	CTT	GCA	CTT	CTC	546
	TCG	TGC	CTG	ACA	ACA	CCA	GCA	TCT	GCC					573

## (2) INFORMATION FOR SEQ ID NO: 142:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs

- 173 -

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGC	CGC	CCC	ATG	GAT	GTA	AAA	TTC	CCG	GGT	78
	GGT	GGT	CAG	ATC	GTT	GGC	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCC	AGG	TTG	GGT	GTG	CGC	GCG	ACT	CGG	AAG	ACT	156
10	TCG	GAG	CGG	TCG	CAA	CCT	CGC	GGC	AGG	CGT	CAG	CCT	ATC	195
	CCC	CAG	GCA	CGT	CGG	TCC	GAG	GGC	AGG	TCC	TGG	GCT	CAG	234
	CCC	GGG	TAC	CCT	TGG	CCT	CTT	TAT	GGC	AAT	GAG	GGC	TGT	273
	GGG	TGG	GCA	GGG	TGG	CTC	CTG	TCC	CCC	CGC	GGA	TCT	CGG	312
	CCA	TCT	TGG	GGC	CAA	AAT	GAT	CCC	CGG	CGT	AGG	TCC	CGC	351
	AAT	CTG	GGT	AAG	GTC	ATC	GAT	ACC	CTG	ACG	TGT	GGC	TTC	390
	GCC	GAC	CTC	ATG	GGA	TAC	ATT	CCG	CTC	GTC	GGC	GCC	CCA	429
15	GTA	GGT	GGC	GTC	GCC	AGG	GCC	TTG	GCG	CAT	GGC	GTC	AGG	468
	GCC	CTG	GAG	GAC	GGA	ATC	AAC	TAT	GCA	ACA	GGG	AAT	CTT	507
	CCT	GGT	TGC	TCC	TTT	TCT	ATC	TTC	CTA	CTT	GCA	CTT	TTC	546
	TCG	TGC	TTG	ACA	ACA	CCG	GCA	TCC	GCT					573

## (2) INFORMATION FOR SEQ ID NO: 143:

20

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

25

(A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z6

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	CGT	39
	AAC	ACC	AAC	CGC	CGC	CCC	ATG	GAC	GTT	AAG	TTC	CCG	GGT	78
	GGT	GGC	CAG	ATC	GTT	GGC	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
30	AGG	GGC	CCC	AGG	TTG	GGT	GTG	CGC	GCG	ACT	AGG	AAG	ACT	156
	TCG	GAG	CGG	TCG	CAA	CCT	CGT	GGG	AGA	CGC	CAG	CCT	ATC	195
	CCC	AAG	GCA	CGT	CGA	TCT	GAG	GGA	AGG	TCC	TGG	GCT	CAG	234
	CCC	GGG	TAT	CCA	TGG	CCT	CTT	TAC	GGT	AAT	GAG	GGT	TGC	273
	GGG	TGG	GCG	GGA	TGG	CTC	CTG	TCA	CCC	CGT	GGC	TCT	CGA	312
	CCG	TCT	TGG	GGT	CCA	AAT	GAT	CCC	CGG	CGA	AGG	TCC	CGC	351
	AAC	TTG	GGT	AAG	GTC	ATC	GAT	ACT	CTA	ACT	TGC	GGT	TTC	390
35	GCC	GAT	CTC	ATG	GGA	TAC	ATC	CCG	CTC	GTA	GGC	GCC	CCC	429
	GTG	GGC	GGC	GTC	GCC	AGG	GCC	CTG	GCA	CAT	GGT	GTT	AGG	468

- 174 -

GCT GTG GAG GAC GGG ATC AAT TAT GCA ACA GGG AAT CTT	507
CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCA CTT CTT	546
TCG TGC CTA ACT GTT CCC ACC TCG GCC	573

## (2) INFORMATION FOR SEQ ID NO: 144:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

10

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z7

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
AAC ACC AAC CGC CGC CCC ATG GAC GTT AAG TTC CCG GGC	78
GGT GGC CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC	117
AGG GGC CCC AGA TTG GGT GTG CGC ACA ACT AGG AAG ACT	156
TCG GAG CGG TCG CAA CCT CGT GGG AGA CGT CAG CCT ATC	195
CCC AAG GCA CGT CGA TCT GAG GGA AGG TCC TGG GCT CAA	234
CCC GGG TAC CCA TGG CCT CTT TAC GGT AAC GAG GGT TGC	273
GGG TGG GCA GGA TGG CTC TTG TCA CCC CGT GGC TCT CGA	312
CCG TCT TGG GGC CCA AAT GAT CCC CGG CGA AGG TCC CGC	351
AAC TTG GGT AAG GTC ATC GAT ACC CTA ACC TGC GGC TTT	390
GCC GAC CTC ATG GGA TAC ATC CCG CTC GTA GGC GCC CCC	429
GTG GGC GGC GTC GCC AGG GCC CTA GCG CAT GGC GTT AGG	468
GCT CTG GAG GAC GGG ATT AAT TAT GCA ACA GGG AAC CTT	507
CCC GGT TGC TCT TTT TCT ATC TTC CTC TTG GCA CTT CTT	546
TCG TGC CTG ACT GTT CCC GCC TCG GCC	573

25

## (2) INFORMATION FOR SEQ ID NO: 145:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK13

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT	39
AAC ACC AAC CGC CGC CCA ATG GAC GTT AAG TTC CCG GGT	78
GGC GGC CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC	117

- 175 -

AGG GGC CCT AGA TTG GGT GTG CGC GCG ACT AGG AAG ACT 156  
 TCG GAG CGG TCG CAA CCT CGT GGG AGG CGC CAG CCT ATC 195  
 CCC AAG GCG CGC CAA CTC GAG GGT AGG TCC TGG GCT CAG 234  
 CCT GGG TAT CCT TGG CCC CTT TAC GGC AAT GAG GGC TGC 273  
 GGG TGG GCG GGA TGG CTC CTG TCA CCC CGT GGC TCT CGG 312  
 CCG TCT TGG GGC CCG AAT GAT CCC CGG CGG AGG TCC CGC 351  
 5 AAC TTG GGT AAG GTC ATC GAT ACC CTA ACT TGC GGC TTC 390  
 GCC GAC CTC ATG GGA TAC ATC CCG GTC GTA GGC GCC CCC 429  
 GTG GGT GGC GTC GCC AGA GCC CTG GCG CAT GGC GTC AGG 468  
 CTT CTG GAG GAC GGG GTC AAT TAT GCA ACA GGG AAT CTT 507  
 CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCA CTG CTC 546  
 TCG TGC CTG ACT GTT CCC GCT TCG GCC 573

10 (2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

20 ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA 39  
 AAC ACC AAC CGC CGC CCA CAG GAC GTT AAG TTC CCG GGC 78  
 GGT GGT CAG ATC GTT GGT GGA GTC TAC TTG TTG CCG CGC 117  
 AGG GGC CCT AGG TTG GGT GTG CGC GCG ACT CGG AAG ACT 156  
 TCA GAA CGG TCG CAA CCC CGT GGG CGG CGC CAG CCT ATT 195  
 CCC AAG GCG CGC CAA CCC ACG GGC CGG TCC TGG GGT CAA 234  
 CCC GGG TAC CCT TGG CCC CTT TAC GCC AAT GAG GGC CTC 273  
 GGG TGG GCA GGG TGG TTG CTC TCC CCC CGA GGC TCT CGG 312  
 25 CCT AAT TGG GGC CCC AAT GAC CCC CGG CGA AAG TCG CGC 351  
 AAT TTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGA TTC 390  
 GCC GAC CTC ATG GGG TAC ATC CCG CTC GTA GGC GGC CCC 429  
 GTT GGG GGC GTC GCA AGG GCC CTT GCA CAT GGT GTG AGG 468  
 GTT CTT GAG GAC GGG GTA AAC TAT GCA ACG GGG AAT TTG 507  
 CCC GGT TGC TCT TTC TCT ATC TTT ATC CTT GCA CTT CTC 546  
 TCG TGC CTG ACC GTC CCG GCC TCT GCA 573

30 (2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 35 (D) TOPOLOGY: linear

- 176 -

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SA5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC	117
	AGG GGC CCT AGA TTG GGT GTG CGC GCG ACT CGG AAG ACT	156
	TCA GAA CGG TCG CAA CCC CGT GGG CGG CGC CAG CCT ATT	195
	CCC AAG GCG CGC CAA CCC ACG GGC CGG TCC TGG GGT CAA	234
	CCC GGG TAC CCT TGG CCC CTT TAC GCC AAT GAG GGC CTC	273
10	GGG TGG GCA GGG TGG TTG CTC TCC CCC CGA GGC TCT CGG	312
	CCT AAT TGG GGC CCC AAT GAC CCC CGG CGA AAA TCG CGC	351
	AAT TTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGA TTC	390
	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTA GGC GGC CCC	429
	GTT GGG GGC GTC GCA AGG GCC CTC GCA CAT GGT GTG AGG	468
	GTT CTT GAG GAC GGG GTA AAC TAT GCA ACA GGG AAT TTG	507
	CCC GGT TGC TCT TTC TCT ATC TTT ATC CTT GCA CTT CTC	546
15	TCG TGC TTG ACC GTC CCA GCC TCT GCA	573

## (2) INFORMATION FOR SEQ ID NO: 148:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 573 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: homosapiens

(C) INDIVIDUAL ISOLATE: SA7

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

25	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC	117
	AGG GGC CCT AGG TTG GGT GTG CGC GCG ACT CGG AAG ACT	156
	TCA GAA CGG TCG CAA CCC CGT GGG CGG CGC CAG CCT ATT	195
	CCC AAG GCG CGC CAA CCC ACG GGC CGG TCC TGG GGT CAA	234
	CCC GGG TAC CCT TGG CCC CTT TAC GCC AAT GAG GGC CTC	273
30	GGG TGG GCA GGG TGG TTG CTC TCC CCC CGA GGC TCT CGG	312
	CCT AAT TGG GGC CCC AAT GAC CCC CGG CGA AAG TCG CGC	351
	AAT TTG GGT AAG GTC ATC GAC ACC CTA ACA TGC GGA TTC	390
	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTA GGC GGC CCC	429
	GTT GGG GGC GTC GCA AGG GCT CTC GCA CAC GGT GTG AGG	468
	GTT CTT GAG GAC GGG GTA AAT TAC GCA ACA GGG AAT CTG	507
	CCC GGT TGC TCT TTC TCT ATC TTT ATC CTT GCA CTT CTC	546
35	TCG TGC CTG ACC GTC CCA GCC TCC GCA	573

- 177 -

## (2) INFORMATION FOR SEQ ID NO: 149:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

10	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC AAC CTC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT CGG AAG ACT	156
	TCG GAA CGG TCG CAA CCC CGT GGG CGG CGC CAG CCT ATT	195
	CCC AAG GCG CGC CAA CCC ACG GGC CGG TCC TGG GGT CAA	234
	CCC GGG TAC CCT TGG CCC CTT TAC GCC AAT GAG GGC CTC	273
15	GGG TGG GCA GGG TGG TTG CTC TCC CCC CGA GGC TCT CGG	312
	CCT AAT TGG GGC CCC AAT GAC CCC CGG CGG AAG TCG CGC	351
	AAT TTG GGT AAG GTC ATC GAT ACC CTA ACG TGC GGA TTC	390
	GCC GAC CTC ATG GGG TAC ATC CCG CTC GTA GGC GGC CCC	429
	GTT GGG GGC GTC GCA AGG GCT CTC GCA CAC GGT GTG AGG	468
	GTT CTT GAG GAC GGG GTA AAC TAC GCA ACA GGG AAT TTG	507
	CCC GGT TGC TCT TTC TCT ATC TTT ATC CTT GCA CTT CTT	546
20	TCC TGT CTG ATC ATC CCG GCC TCT GCA	573

## (2) INFORMATION FOR SEQ ID NO: 150:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA3

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

30	ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA AGA	39
	AAC ACC AAC CGC CGC CCA CAG GAC GTC AAG TTC CCG GGC	78
	GGT GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC	117
	AGG GGC CCC AGG TTG GGT GTG CGC GCG ACT CGG AAG ACT	156
	TCA GAA CGG TCG CAA CCC CGT GGA CGG CGC CAG CCT ATT	195
	CCC AAG GCT CGC CAG CCC ACG GGC CGG TCC TGG GGT CAA	234
35	CCC GGG TAC CCT TGG CCC CTT TAC GCC AAT GAG GGC CTC	273
	GAG TGG GCA GGG TGG TTG CTC TCC CCC CGA GGC TCT CGG	312

- 178 -

	CCT	AGT	TGG	GGC	CCC	AAC	GAC	CCC	CGG	CGG	AAA	TCG	CGC	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTA	ACG	TGC	GGA	TTC	390
	GCC	GAT	CTC	ATG	GGG	TAC	ATC	CCG	CTC	GTA	GGC	GGC	CCC	429
	GTT	GGG	GGC	GTC	GCA	AGG	GCT	CTC	GCA	CAT	GGT	GTG	AGG	468
	GTT	CTT	GAG	GAC	GGG	GTA	AAC	TAC	GCA	ACA	GGG	AAT	TTA	507
	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTT	ATC	CTT	GCA	CTT	CTT	546
5	TCA	TGC	CTG	ACC	GTC	CCG	GCC	TCT	GCA					573

## (2) INFORMATION FOR SEQ ID NO: 151:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 573 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: SA13
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	AGA	39
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	78
	GGT	GGT	CAG	ATC	GTT	GGT	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCT	AGG	TTG	GGT	GTG	CGC	GCA	ACT	CGG	AAG	ACT	156
	TCA	GAA	CGG	TCG	CAA	CCC	CGT	GGA	CGG	CGT	CAG	CCT	ATC	195
20	CCC	AAG	GCG	CGC	CAG	CCC	ACG	GGC	CGG	TCC	TGG	GGT	CAA	234
	CCC	GGG	TAC	CCT	TGG	CCC	CTT	TAT	GCC	AAT	GAG	GGC	CTC	273
	GGG	TGG	GCA	GGG	TGG	TTG	CTC	TCC	CCC	CGA	GGC	TCT	CGG	312
	CCT	AAT	TGG	GGC	CCC	AAT	GAC	CCC	CGG	CGG	AAA	TCG	CGC	351
	AAC	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTG	ACG	TGC	GGA	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATC	CCG	CTC	GTA	GGC	GGC	CCC	429
	GTT	GGG	GGC	GTC	GCA	AGG	GCT	CTC	GCA	CAC	GGT	GTG	AGG	468
25	GTC	CTT	GAG	GAC	GGG	GTA	AAC	TAT	GCA	ACA	GGG	AAT	TTA	507
	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTT	ATC	CTT	GCA	CTT	CTT	546
	TCA	TGC	CTG	ACT	GTC	CCG	ACC	TCT	GCC					573

## (2) INFORMATION FOR SEQ ID NO: 152:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 573 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: SA6



- 179 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	CAA	AGA	39
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	78
	GGT	GGT	CAG	ATC	GTT	GGT	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCT	CGT	ATG	GGT	GTG	CGC	GCG	ACT	CGG	AAG	ACT	156
5	TCG	GAA	CGG	TCG	CAA	CCC	CGT	GGA	CGG	CGT	CAG	CCT	ATT	195
	CCC	AAG	GCG	CGC	CAA	TCC	GCG	GGT	CGG	TCC	TGG	GGT	CAA	234
	CCC	GGG	TAC	CCT	TGG	CCC	CTT	TAC	GCC	AAT	GAG	GGC	CTC	273
	GGG	TGG	GCA	GGG	TGG	TTG	CTC	TCC	CCC	CGA	GGC	TCT	CGG	312
	CCT	AAT	TGG	GGC	CCC	AAT	GAC	CCC	CGG	CGA	AAA	TCG	CGC	351
	AAT	TTG	GGT	AAG	GTC	ATC	GAT	ACC	CTA	ACG	TGC	GGA	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATC	CCG	CTC	GTA	GGC	GGC	CCC	429
10	GTT	GGG	GGC	GTC	GCA	AGG	GCT	CTC	GCA	CAC	GGT	GTG	AGG	468
	GTT	CTT	GAG	GAC	GGG	GTA	AAC	TAT	GCA	ACA	GGG	AAT	TTG	507
	CCC	GGT	TGC	TCT	TTC	TCT	ATC	TTT	GTC	CTT	GCA	CTT	CTC	546
	TCG	TGC	CTA	ACC	GTC	CCT	GCC	TCT	GCA					573

## (2) INFORMATION FOR SEQ ID NO: 153:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 20 (C) INDIVIDUAL ISOLATE: SA11

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

	ATG	AGC	ACG	AAT	CCT	AAA	CCT	CAA	AGA	AAA	ACC	AAA	AGA	39
	AAC	ACC	AAC	CGC	CGC	CCA	CAG	GAC	GTC	AAG	TTC	CCG	GGC	78
	GGT	GGT	CAG	ATC	GTT	GGT	GGA	GTT	TAC	TTG	TTG	CCG	CGC	117
	AGG	GGC	CCT	AGG	TTG	GGT	GTG	CGC	GCG	ACT	CGG	AAG	ACT	156
25	TCA	GAA	CGG	TCG	CAA	CCC	CGT	GGG	CGG	CGT	CAG	CCT	ATT	195
	CCC	AAG	GCG	CGC	CAA	CCC	ACG	GGC	CGG	TCC	TGG	GGT	CAA	234
	CCC	GGG	TAC	CCT	TGG	CCC	TTT	TAC	GCC	AAT	GAG	GGC	CTC	273
	GGG	TGG	GCA	GGG	TGG	CTG	CTC	TCC	CCT	CGA	GGC	TCT	CGG	312
	CCT	AAC	TGG	GGC	CCC	AAT	GAC	CCC	CGG	CGA	AGA	TCG	CGC	351
	AAT	TTG	GGC	AAG	GTC	ATC	GAT	ACC	CTA	ACG	TGC	GGA	TTC	390
	GCC	GAC	CTC	ATG	GGG	TAC	ATC	CCG	CTC	GTA	GGC	GGC	CCC	429
30	GTT	GGG	GGC	GTC	GCA	AGG	GCC	CTC	GCA	CAC	GGT	GTG	AGA	468
	GCT	CTT	GAG	GAC	GGG	GTA	AAT	TAT	GCA	ACA	GGG	AAT	CTT	507
	CCC	GGT	TGC	TCT	TTC	TCC	ATC	TTT	ATC	CTT	GCA	CTT	CTC	546
	TCG	TGC	TTG	ACC	GTC	CCG	GCC	ACT	GCA					573

## (2) INFORMATION FOR SEQ ID NO: 154:

35 (i) SEQUENCE CHARACTERISTICS:

- 180 -

(A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

	ATG AGC ACA CTT CCA AAA CCC CAA AGA AAA ACC AAA AGA	39
	AAC ACC AAC CGT CGC CCA ACG GAC GTC AAG TTC CCG GGT	78
10	GGC GGT CAG ATC GTT GGC GGA GTT TAC TTG TTG CCG CGC	117
	AGG GGC CCC CGG TTG GGT GTG CGC GCG ACG AGA AAG ACT	156
	TCC GAG CGA TCC CAG CCC AGA GGC AGG CGC CAA CCT ATA	195
	CCA AAG GCG CGC CAG CCC CAG GGC AGG CAC TGG GCT CAG	234
	CCC GGA TAC CCT TGG CCT CTT TAT GGA AAC GAG GGC TGT	273
	GGG TGG GCA GGT TGG CTC CTG TCC CCC CGC GGC TCC CGG	312
	CCA CAT TGG GGC CCC AAT GAC CCC CGG CGT CGA TCC CGG	351
	AAT TTG GGT AAG GTC ATC GAT ACC CTA ACG TGT GGG TTC	390
15	GCC GAT CTC ATG GGG TAC ATT CCC GTC GTG GGC GCG CCT	429
	TTG GGC GGC GTC GCG GCT GCG CTC GCA CAT GGC GTG AGG	468
	GCA ATC GAG GAC GGG ATC AAT TAT GCA ACA GGG AAT CTC	507
	CCC GGT TGC TCT TTC TCT ATC TTC CTT TTG GCA CTA CTC	546
	TCG TGC CTC ACA ACG CCA GCT TCG GCT	573

20 (2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

25 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

	Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn	
	1 5 10	
30	Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly	
	15 20 25	
	Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro	
	30 35 40	
	Arg Leu Gly Val Arg Ala Pro Arg Lys Thr Ser Glu Arg Ser	
	45 50 55	
35	Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg	
	60 65 70	

- 181 -

Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 156:

- 15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: unknown
- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Il Pro Leu Val  
 130 135 140

- 182 -

Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 5 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 25 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 30 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190  
 35

- 183 -

## (2) INFORMATION FOR SEQ ID NO: 158:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 159:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- 184 -

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S18

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

```

5  Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
   1      5      10
  Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
   15      20      25
  Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro
   30      35      40
  Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
   45      50      55
10  Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
   60      65      70
  Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro
   75      80
  Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu
   85      90      95
  Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
  100      105      110
15  Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
   115      120      125
  Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
   130      135      140
  Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly
   145      150
20  Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
   155      160      165
  Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
   170      175      180
  Ser Cys Leu Thr Val Pro Ala Ser Ala
   185      190

```

25 (2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

30 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DR4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

```

35  Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
   1      5      10

```

- 185 -

Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 161:

20

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

25

- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: SA10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80

- 186 -

Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 5 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Pro Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 10 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 20 (C) INDIVIDUAL ISOLATE: S45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

Met Ser Thr Asn Pro Lys Pro Gln Arg Ala Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly His Pro Trp Pro  
 75 80  
 30 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 35 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140



- 187 -

Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 163:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: D1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

- 188 -

## (2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US6

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

10	Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn
	1				5					10				
	Thr	Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly
	15				20					25				
	Gln	Ile	Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro
	30				35					40				
	Arg	Leu	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser
	45				50					55				
15	Gln	Pro	Arg	Gly	Arg	Arg	Gln	Pro	Ile	Pro	Lys	Ala	Arg	Arg
	60				65					70				
	Pro	Glu	Gly	Arg	Ala	Trp	Ala	Gln	Pro	Gly	Tyr	Pro	Trp	Pro
	75				80					85				
	Leu	Tyr	Gly	Asn	Glu	Gly	Met	Gly	Trp	Ala	Gly	Trp	Leu	Leu
	90				95					100				
20	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	Gly	Pro	Thr	Asp	Pro
	105				110					115				
	Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys	Val	Ile	Asp	Thr	Leu
	120				125					130				
	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Leu	Val
	135				140					145				
	Gly	Ala	Pro	Leu	Gly	Gly	Ala	Ala	Arg	Ala	Leu	Ala	His	Gly
	150				155					160				
25	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr	Ala	Thr	Gly	Asn
	165				170					175				
	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile	Phe	Leu	Leu	Ala	Leu	Leu
	180				185					190				
	Ser	Cys	Leu	Thr	Ile	Pro	Ala	Ser	Ala					

## 30 (2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

35 (vi) ORIGINAL SOURCE:

- 189 -

(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: P10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

```

5  Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
   1      5      10
  Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
   15      20      25
  Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro
   30      35      40
  Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
   45      50      55
10  Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
   60      65      70
  Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro
   75      80
  Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu
   85      90      95
  Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
  100      105      110
15  Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
   115      120      125
  Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
   130      135      140
  Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly
   145      150
  Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
  20  155      160      165
  Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
  170      175      180
  Ser Cys Leu Thr Ile Pro Ala Ser Ala
   185      190

```

25 (2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

30 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

```

35  Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
   1      5      10

```

- 190 -

Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 167:

20

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

25

- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: T10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80

- 191 -

Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 5 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 10 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown  
 15

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SW2  
 20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:  
 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 30 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 35

- 192 -

Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
                                   145                                  150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155                                  160                                  165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
           170                                  175                                  180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
                                   185                                  190

## (2) INFORMATION FOR SEQ ID NO: 169:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: IND3

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
   1                                  5                                  10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
   15                                  20                                  25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
   30                                  35                                  40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
   45                                  50                                  55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
   60                                  65                                  70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
   75                                  80  
 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
   85                                  90                                  95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
   100                                  105                                  110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
   115                                  120                                  125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
   130                                  135                                  140  
 Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly  
   145                                  150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155                                  160                                  165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
   170                                  175                                  180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
                                   185                                  190

- 193 -

## (2) INFORMATION FOR SEQ ID NO: 170:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: IND8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

```

10 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
    1      5      10
    Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
    15      20      25
    Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro
    30      35      40
    Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
    45      50      55
15 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
    60      65      70
    Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly His Pro Trp Pro
    75      80
    Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu
    85      90      95
    Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
    100      105      110
20 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
    115      120      125
    Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
    130      135      140
    Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly
    145      150
25 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
    155      160      165
    Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
    170      175      180
    Ser Cys Leu Thr Val Pro Ala Ser Ala
    185      190

```

## 30 (2) INFORMATION FOR SEQ ID NO: 171:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- 194 -

(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: S9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

```

5  Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
   1      5      10
  Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
  15      20      25
  Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro
   30      35      40
  Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
   45      50      55
10  Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg His
   60      65      70
  Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro
   75      80
  Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu
   85      90      95
  Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro
  100      105      110
15  Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
   115      120      125
  Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
   130      135      140
  Gly Ala Pro Leu Gly Gly Ala Ala Arg Ala Leu Ala His Gly
   145      150
  Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
  155      160      165
20  Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
   170      175      180
  Ser Cys Leu Thr Ile Pro Ala Ser Ala
   185      190

```

(2) INFORMATION FOR SEQ ID NO: 172:

25

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 191 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

30

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: homosapiens  
(C) INDIVIDUAL ISOLATE: HK3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

```

35  Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
   1      5      10
  Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
   15      20      25

```



- 195 -

Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 5 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 10 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 15 Ser Cys Leu Thr Thr Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown  
 20

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK5  
 25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:  
 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 30 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 35

- 196 -

Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro His Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 5 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Ile Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 10 Ser Cys Leu Thr Thr Pro Val Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 174:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK4

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 25 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 30 Leu Tyr Gly Asn Glu Gly Met Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 35 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150

- 197 -

Val Arg Val Val Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

5

## (2) INFORMATION FOR SEQ ID NO: 175:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: P8

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

15

Met Ser Thr Thr Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Ser Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly His Pro Trp Pro  
 75 80  
 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Gly Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Val Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

35

## (2) INFORMATION FOR SEQ ID NO: 176:

- 198 -

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

5

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T3

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

10 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 15 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Pro Glu Gly Arg Ala Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asp Glu Gly Met Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Thr Asp Pro  
 100 105 110  
 20 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 25 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Ile Pro Ala Ser Ala  
 185 190

- (2) INFORMATION FOR SEQ ID NO: 177:

30

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

35

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens

- 199 -

## (C) INDIVIDUAL ISOLATE: T4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

```

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
 1      5      10
5 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
 15      20      25
Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro
 30      35      40
Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
 45      50      55
Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg
 60      65      70
10 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro
 75      80
Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu
 85      90      95
Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro
100      105      110
Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu
115      120      125
15 Thr Cys Ser Leu Ala Asp Leu Met Gly Tyr Val Pro Val Val
 130      135      140
Gly Gly Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly
 145      150
Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn
155      160      165
20 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
 170      175      180
Ser Cys Ile Thr Ile Pro Val Ser Ala
 185      190

```

(2) INFORMATION FOR SEQ ID NO: 178:

```

25 (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 191 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS: unknown
      (D) TOPOLOGY: unknown

```

```

      (vi) ORIGINAL SOURCE:
            (A) ORGANISM: homosapiens
            (C) INDIVIDUAL ISOLATE: US10
30

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

```

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
 1      5      10
35 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
 15      20      25

```

- 200 -

0 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
    30                  35                  40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
           45                  50                  55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg  
           60                  65                  70  
 Pro Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro  
           75                  80  
 5 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
    85                  90                  95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
    100                  105                  110  
 Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu  
           115                  120                  125  
 10 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
           130                  135                  140  
 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
           145                  150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
    155                  160                  165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
    170                  175                  180  
 15 Ser Cys Ile Thr Ile Pro Val Ser Ala  
           185                  190

## (2) INFORMATION FOR SEQ ID NO: 179:

## (i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

25

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T9

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

30 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Ile Arg Asn  
    1                  5                  10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
    15                  20                  25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
    30                  35                  40  
 Arg Leu Gly Val Arg Thr Thr Arg Lys Thr Ser Glu Arg Ser  
           45                  50                  55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg  
           60                  65                  70  
 35 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro  
           75                  80

- 201 -

Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Ser Asp Pro  
 100 105 110  
 Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Ile Thr Thr Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 180:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

Met Ser Thr Ile Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg  
 60 65 70  
 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg His Arg Ser Arg Asn Val Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Il Pro Val Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150

- 202 -

Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Ile Thr Ile Pro Val Ser Ala  
 185 190

5

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: T8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

15

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg  
 60 65 70  
 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Thr Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg His Arg Ser Arg Asn Leu Gly Arg Val Ile Asp Thr Ile  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
 130 135 140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Phe Thr Val Pro Val Ser Ala  
 185 190

35



- 203 -

## (2) INFORMATION FOR SEQ ID NO: 182:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: US1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

10	Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn
	1				5					10				
	Thr	Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly
	15					20				25				
	Gln	Ile	Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro
		30				35					40			
	Arg	Leu	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser
		45				50					55			
15	Gln	Pro	Arg	Gly	Arg	Arg	Gln	Pro	Ile	Pro	Lys	Asp	Arg	Arg
			60				65						70	
	Ser	Thr	Gly	Lys	Ser	Trp	Gly	Lys	Pro	Gly	Tyr	Pro	Trp	Pro
				75					80					
	Leu	Tyr	Gly	Asn	Glu	Gly	Cys	Gly	Trp	Ala	Gly	Trp	Leu	Leu
	85			90					95					
	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Thr	Trp	Gly	Pro	Thr	Asp	Pro
	100					105					110			
20	Arg	His	Arg	Ser	Arg	Asn	Leu	Gly	Lys	Val	Ile	Asp	Thr	Ile
		115				120					125			
	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Val	Val
		130				135							140	
	Gly	Ala	Pro	Val	Gly	Gly	Val	Ala	Arg	Ala	Leu	Ala	His	Gly
				145					150					
	Val	Arg	Val	Leu	Glu	Asp	Gly	Ile	Asn	Tyr	Ala	Thr	Gly	Asn
25	155				160					165				
	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile	Phe	Leu	Leu	Ala	Leu	Leu
		170				175						180		
	Ser	Cys	Ala	Thr	Val	Pro	Val	Ser	Ala					
		185						190						

## (2) INFORMATION FOR SEQ ID NO: 183:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens

- 204 -

## (C) INDIVIDUAL ISOLATE: DK11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

```

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
 1      5      10
5 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
 15      20      25
Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro
 30      35      40
Arg Leu Gly Val Arg Thr Thr Arg Lys Thr Ser Glu Arg Ser
 45      50      55
Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg
 60      65      70
10 Ser Thr Gly Lys Pro Trp Gly Lys Pro Gly Tyr Pro Trp Pro
 75      80
Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu
 85      90      95
Ser Pro Arg Gly Ser His Pro Asn Trp Gly Pro Thr Asp Pro
100      105      110
Arg His Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Ile
15      115      120      125
Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val
 130      135      140
Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly
 145      150
Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn
155      160      165
20 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu
 170      175      180
Ser Cys Cys Thr Val Pro Val Ser Ala
 185      190

```

(2) INFORMATION FOR SEQ ID NO: 184:

```

25 (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 191 amino acids
    (B) TYPE: amino acid
    (C) STRANDEDNESS: unknown
    (D) TOPOLOGY: unknown

```

```

    (vi) ORIGINAL SOURCE:
    (A) ORGANISM: homosapiens
30 (C) INDIVIDUAL ISOLATE: SW3

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

```

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
 1      5      10
35 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
 15      20      25

```

- 205 -

0 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
    30                   35                   40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
           45                   50                   55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg  
           60                   65                   70  
 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro  
                    75                   80  
 5 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
    85                   90                   95  
 Ser Pro Arg Gly Ser His Pro Asn Trp Gly Pro Thr Asp Pro  
    100                   105                   110  
 Arg His Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Ile  
           115                   120                   125  
 10 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
           130                   135                   140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
           145                   150  
 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
    155                   160                   165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
    170                   175                   180  
 15 Ser Cys Phe Thr Val Pro Val Ser Ala  
           185                   190

## (2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown  
 20  
 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK8  
 25  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:  
 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
    1                   5                   10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
    15                   20                   25  
 30 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
    30                   35                   40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Ser Ser Glu Arg Ser  
           45                   50                   55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg  
           60                   65                   70  
 Ser Thr Gly Lys Ser Trp Gly Lys Pro Gly Tyr Pro Trp Pro  
           75                   80  
 35 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
    85                   90                   95

- 206 -

Ser Pro Arg Gly Ser Arg Pro Thr Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg His Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Ile  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
 130 135 140  
 5 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Cys Thr Val Pro Val Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 186:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: S83

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro  
 30 35 40  
 25 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Asp Arg Arg  
 60 65 70  
 Thr Thr Gly Lys Ser Trp Gly Arg Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 30 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro  
 100 105 110  
 Arg His Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
 130 135 140  
 35 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150

- 207 -

Val Arg Val Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Ile Ser Val Pro Val Ser Ala  
 185 190

5

## (2) INFORMATION FOR SEQ ID NO: 187:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

15

Met Ser Thr Leu Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Ile Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Val Leu Pro Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Ala Leu Glu Asp Gly Ile Asn Phe Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe  
 170 175 180  
 Ser Cys Leu Ile His Pro Ala Ala Ser  
 185 190

35

- 208 -

## (2) INFORMATION FOR SEQ ID NO: 188:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: S52

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

```

10 Met Ser Thr Leu Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
    1      5      10
    Thr Ile Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
    15      20      25
    Gln Ile Val Gly Gly Val Tyr Val Leu Pro Arg Arg Gly Pro
    30      35      40
    Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
    45      50      55
15 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
    60      65      70
    Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro
    75      80
    Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu
    85      90      95
    Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro
    100      105      110
20 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
    115      120      125
    Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
    130      135      140
    Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly
    145      150
    Val Arg Ala Leu Glu Asp Gly Ile Asn Phe Ala Thr Gly Asn
    155      160      165
25 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe
    170      175      180
    Ser Cys Leu Val His Pro Ala Ala Ser
    185      190

```

## (2) INFORMATION FOR SEQ ID NO: 189:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens

- 209 -

## (C) INDIVIDUAL ISOLATE: S2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

```

Met Ser Thr Leu Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
 1      5      10
5 Thr Ile Arg Arg Pro Gln Asp Ile Lys Phe Pro Gly Gly Gly
 15      20      25
Gln Ile Val Gly Gly Val Tyr Val Leu Pro Arg Gly Pro
 30      35      40
Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser
 45      50      55
Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg
 60      65      70
10 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro
 75      80
Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu
 85      90      95
Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro
100      105      110
Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu
115      120      125
15 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val
 130      135      140
Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly
 145      150
Val Arg Ala Leu Glu Asp Gly Ile Asn Phe Ala Thr Gly Asn
155      160      165
20 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe
 170      175      180
Ser Cys Leu Ile His Pro Ala Ala Ser
 185      190

```

## (2) INFORMATION FOR SEQ ID NO: 190:

```

25 (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 191 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS: unknown
      (D) TOPOLOGY: unknown

```

```

30 (vi) ORIGINAL SOURCE:
      (A) ORGANISM: homosapiens
      (C) INDIVIDUAL ISOLATE: DK12

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

```

Met Ser Thr Leu Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
 1      5      10
35 Thr Ile Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly
 15      20      25

```

- 210 -

0 Gln Ile Val Gly Gly Val Tyr Val Leu Pro Arg Arg Gly Pro  
    30                   35                   40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
           45                   50                   55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
           60                   65                   70  
 5 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
           75                   80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
    85                   90                   95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
    100                   105                   110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
           115                   120                   125  
 10 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
           130                   135                   140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
           145                   150  
 Val Arg Ala Leu Glu Asp Gly Ile Asn Phe Ala Thr Gly Asn  
    155                   160                   165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Phe  
    170                   175                   180  
 15 Ser Cys Leu Ile His Pro Ala Ala Ser  
           185                   190

## (2) INFORMATION FOR SEQ ID NO: 191:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z4

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
    1                   5                   10  
 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly  
    15                   20                   25  
 30 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
           30                   35                   40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
           45                   50                   55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
           60                   65                   70  
 Pro Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
           75                   80  
 35 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
    85                   90                   95



- 211 -

Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Ile Val  
 130 135 140  
 5 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

10

## (2) INFORMATION FOR SEQ ID NO: 192:

## (i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z8

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

20

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 25 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 30 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150

35

- 212 -

Val Arg Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

5

## (2) INFORMATION FOR SEQ ID NO: 193:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z1

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

15 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Ala Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 20 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 25 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 30 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Thr Pro Ala Ser Ala  
 185 190

35

## (2) INFORMATION FOR SEQ ID NO: 194:

- 213 -

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

5 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

	Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn
	1				5					10				
10	Thr	Asn	Arg	Arg	Pro	Met	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly
	15				20					25				
	Gln	Ile	Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro
	30				35					40				
	Arg	Leu	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser
	45				50					55				
	Gln	Pro	Arg	Gly	Arg	Arg	Gln	Pro	Ile	Pro	Gln	Ala	Arg	Arg
15				60					65					70
	Ser	Glu	Gly	Arg	Ser	Trp	Ala	Gln	Pro	Gly	Tyr	Pro	Trp	Pro
				75					80					
	Leu	Tyr	Gly	Asn	Glu	Gly	Cys	Gly	Trp	Ala	Gly	Trp	Leu	Leu
	85				90				95					
	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Ser	Trp	Gly	Gln	Asn	Asp	Pro
	100				105					110				
20	Arg	Arg	Arg	Ser	Arg	Asn	Leu	Gly	Lys	Val	Ile	Asp	Thr	Leu
				115				120				125		
	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Leu	Val
				130				135						140
	Gly	Ala	Pro	Val	Gly	Gly	Val	Ala	Arg	Ala	Leu	Ala	His	Gly
				145				150						
	Val	Arg	Ala	Leu	Glu	Asp	Gly	Ile	Asn	Tyr	Ala	Thr	Gly	Asn
	155				160					165				
25	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile	Phe	Leu	Leu	Ala	Leu	Phe
	170				175					180				
	Ser	Cys	Leu	Thr	Thr	Pro	Ala	Ser	Ala					
			185					190						

(2) INFORMATION FOR SEQ ID NO: 195:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z6

- 214 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
 60 65 70  
 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Ala Val Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Thr Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 196:

## (i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

30 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: Z7

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 35 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro  
 30 35 40

- 215 -

Arg Leu Gly Val Arg Thr Thr Arg Lys Thr Ser Glu Arg Ser  
                   45                                  50                                  55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg  
                   60                                  65                                  70  
 Ser Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
                   75                                  80  
 5 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
    85                                  90                                  95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
    100                                  105                                  110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
                   115                                  120                                  125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
                   130                                  135                                  140  
 10 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
                   145                                  150  
 Val Arg Ala Leu Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
    155                                  160                                  165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
    170                                  175                                  180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
                   185                                  190  
 15

## (2) INFORMATION FOR SEQ ID NO: 197:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown  
 20

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: DK13

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
    1                                  5                                  10  
 Thr Asn Arg Arg Pro Met Asp Val Lys Phe Pro Gly Gly Gly  
    15                                  20                                  25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
    30                                  35                                  40  
 30 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
                   45                                  50                                  55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
                   60                                  65                                  70  
 Leu Glu Gly Arg Ser Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
                   75                                  80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
    85                                  90                                  95  
 35

- 216 -

Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
 130 135 140  
 Gly Ala Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Leu Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

10

(2) INFORMATION FOR SEQ ID NO: 198:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

15

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

20 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 25 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro  
 100 105 110  
 30 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 35 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165

- 217 -

Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

## 5 (2) INFORMATION FOR SEQ ID NO: 199:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SAS

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 15 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 20 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro  
 100 105 110  
 25 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 30 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 200:

35

- (i) SEQUENCE CHARACTERISTICS:

- 218 -

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: SA7

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

	Met	Ser	Thr	Asn	Pro	Lys	Pro	Gln	Arg	Lys	Thr	Lys	Arg	Asn
	1				5					10				
10	Thr	Asn	Arg	Arg	Pro	Gln	Asp	Val	Lys	Phe	Pro	Gly	Gly	Gly
	15				20					25				
	Gln	Ile	Val	Gly	Gly	Val	Tyr	Leu	Leu	Pro	Arg	Arg	Gly	Pro
	30					35					40			
	Arg	Leu	Gly	Val	Arg	Ala	Thr	Arg	Lys	Thr	Ser	Glu	Arg	Ser
	45					50					55			
	Gln	Pro	Arg	Gly	Arg	Arg	Gln	Pro	Ile	Pro	Lys	Ala	Arg	Gln
	60					65					70			
15	Pro	Thr	Gly	Arg	Ser	Trp	Gly	Gln	Pro	Gly	Tyr	Pro	Trp	Pro
	75					80								
	Leu	Tyr	Ala	Asn	Glu	Gly	Leu	Gly	Trp	Ala	Gly	Trp	Leu	Leu
	85				90				95					
	Ser	Pro	Arg	Gly	Ser	Arg	Pro	Asn	Trp	Gly	Pro	Asn	Asp	Pro
	100					105					110			
	Arg	Arg	Lys	Ser	Arg	Asn	Leu	Gly	Lys	Val	Ile	Asp	Thr	Leu
	115					120					125			
20	Thr	Cys	Gly	Phe	Ala	Asp	Leu	Met	Gly	Tyr	Ile	Pro	Leu	Val
	130					135					140			
	Gly	Gly	Pro	Val	Gly	Gly	Val	Ala	Arg	Ala	Leu	Ala	His	Gly
	145					150								
	Val	Arg	Val	Leu	Glu	Asp	Gly	Val	Asn	Tyr	Ala	Thr	Gly	Asn
	155				160					165				
	Leu	Pro	Gly	Cys	Ser	Phe	Ser	Ile	Phe	Ile	Leu	Ala	Leu	Leu
	170					175					180			
25	Ser	Cys	Leu	Thr	Val	Pro	Ala	Ser	Ala					
	185					190								

## (2) INFORMATION FOR SEQ ID NO: 201:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 191 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens
- (C) INDIVIDUAL ISOLATE: SA1



- 219 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Leu Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 5 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 10 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 15 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu  
 170 175 180  
 20 Ser Cys Leu Ile Ile Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 202:

## (i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA3

## 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 35 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40

- 220 -

Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
                   45                                  50                                  55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
                           60                                  65                                  70  
 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro  
                                   75                                  80  
 Leu Tyr Ala Asn Glu Gly Leu Glu Trp Ala Gly Trp Leu Leu  
 5                   85                                  90                                  95  
 Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Asn Asp Pro  
                   100                                  105                                  110  
 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
                   115                                  120                                  125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
                           130                                  135                                  140  
 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 10                                   145                                  150  
 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
                   155                                  160                                  165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu  
                   170                                  175                                  180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
                           185                                  190

15

## (2) INFORMATION FOR SEQ ID NO: 203:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 191 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

20

- (vi) ORIGINAL SOURCE:
- (A) ORGANISM: homosapiens
  - (C) INDIVIDUAL ISOLATE: SA13

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

25

Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
   1                                  5                                  10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
   15                                  20                                  25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
   30                                  35                                  40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 30                   45                                  50                                  55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
                           60                                  65                                  70  
 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro  
                                   75                                  80  
 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
   85                                  90                                  95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro  
 35                   100                                  105                                  110

- 221 -

Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 5 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Thr Ser Ala  
 185 190

10 (2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

15 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

20 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Gln Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Gly Pro  
 30 35 40  
 Arg Met Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 25 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 Ser Ala Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro  
 100 105 110  
 30 Arg Arg Lys Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 35 Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165

- 222 -

Leu Pro Gly Cys Ser Phe Ser Ile Phe Val Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Ser Ala  
 185 190

## 5 (2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10 (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: SA11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

15 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 Thr Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 20 Pro Thr Gly Arg Ser Trp Gly Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Phe Tyr Ala Asn Glu Gly Leu Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro Asn Trp Gly Pro Asn Asp Pro  
 100 105 110  
 25 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val  
 130 135 140  
 Gly Gly Pro Val Gly Gly Val Ala Arg Ala Leu Ala His Gly  
 145 150  
 Val Arg Ala Leu Glu Asp Gly Val Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 30 Leu Pro Gly Cys Ser Phe Ser Ile Phe Ile Leu Ala Leu Leu  
 170 175 180  
 Ser Cys Leu Thr Val Pro Ala Thr Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO: 206:

35 (i) SEQUENCE CHARACTERISTICS:

- 223 -

- (A) LENGTH: 191 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (vi) ORIGINAL SOURCE:

- 5 (A) ORGANISM: homosapiens  
 (C) INDIVIDUAL ISOLATE: HK2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

Met Ser Thr Leu Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn  
 1 5 10  
 10 Thr Asn Arg Arg Pro Thr Asp Val Lys Phe Pro Gly Gly Gly  
 15 20 25  
 Gln Ile Val Gly Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro  
 30 35 40  
 Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu Arg Ser  
 45 50 55  
 Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Gln  
 60 65 70  
 15 Pro Gln Gly Arg His Trp Ala Gln Pro Gly Tyr Pro Trp Pro  
 75 80  
 Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu  
 85 90 95  
 Ser Pro Arg Gly Ser Arg Pro His Trp Gly Pro Asn Asp Pro  
 100 105 110  
 Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu  
 115 120 125  
 20 Thr Cys Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Val Val  
 130 135 140  
 Gly Ala Pro Leu Gly Gly Val Ala Ala Ala Leu Ala His Gly  
 145 150  
 Val Arg Ala Ile Glu Asp Gly Ile Asn Tyr Ala Thr Gly Asn  
 155 160 165  
 Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu  
 170 175 180  
 25 Ser Cys Leu Thr Thr Pro Ala Ser Ala  
 185 190

## (2) INFORMATION FOR SEQ ID NO:207:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 40 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:207:

35 GCGTCCGGGT TCTGGAAGAC GGC GTGA ACT ATGCAACAGG

40

- 224 -

## (2) INFORMATION FOR SEQ ID NO:208:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:208:

AGGCTTTTCAT TGCAGTTCAA GGCCGTGCTA TTGATGTGCC

40

## (2) INFORMATION FOR SEQ ID NO:209:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:209:

AAGACGGCGT GAACTATGCA ACAGGGAACC TTCCTGGTTG

40

## (2) INFORMATION FOR SEQ ID NO:210:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:210:

AGTTCAAGGC CGTGCTATTG ATGTGCCAAC TGCCGTTGGT

40

## (2) INFORMATION FOR SEQ ID NO:211:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:211:

AAGACGGCGT GAATTCTGCA ACAGGGAACC TTCCTGGTTG

40

- 225 -

## (2) INFORMATION FOR SEQ ID NO:212:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:212:

AGTTCAAGGC CGTGAATTC ATGTGCCAAC TGCCGTTGGT 40

## (2) INFORMATION FOR SEQ ID NO:213:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:213:

ARCTYCGACG TYACATCGAY CTGCTYGTYG GRAGYGCCAC CC 42

## (2) INFORMATION FOR SEQ ID NO:214:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:214:

RCARGCCRTC TTGGAYATGA TCGCTGGWGC Y 31

## (2) INFORMATION FOR SEQ ID NO:215:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:215:

CRATACGACR YCAYGTCGAY TTGCTCGTTG GGGCGGCTRY YT 42

## (2) INFORMATION FOR SEQ ID NO:216:

- 226 -

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:  
RCAAGCTRTC RTGGAYRTGG TRRCRGGRGC C 31
- 10 (2) INFORMATION FOR SEQ ID NO:217:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:  
TTGCGGACKC ACATYGACAT GGTGTGATG TCCGCCACGC 40
- (2) INFORMATION FOR SEQ ID NO:218:
- 20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:218:  
25 GATGCGCGTT CCCGAGGTCA TCWTAGACAT CRTYRGCGGR GCD 43
- (2) INFORMATION FOR SEQ ID NO:219:
- 30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 54 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:  
AATGGCACCY TGCRCCTGCTG GATACAAGTR ACACCTAATG TGGCTGTGAA 50  
ACAC 54
- 35 (2) INFORMATION FOR SEQ ID NO:220:



- 227 -

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:  
TGARCTAGYC CTYSARGTYG TCTTCGGYGG Y 31
- (2) INFORMATION FOR SEQ ID NO:221:
- 10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 54 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:221:  
15 GCCAACGTCT CTCGATGTTG GGTGCCGGTT GCCCCCAATC TCGCCATAAG 50  
TCAA 54
- (2) INFORMATION FOR SEQ ID NO:222:
- 20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 46 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:222:  
25 AAGGGCCTGC GAGCACACAT CGATATCATC GTGATGTCTG CTACGG 46
- 
- (2) INFORMATION FOR SEQ ID NO:223:
- 30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 45 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:223:  
TTGGTGCGCA TCCCGGAAGT CATCTTGGAT ATTGTTACAG GAGGT 45
- 35 (2) INFORMATION FOR SEQ ID NO:224:

- 228 -

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:  
AGTCAGGTAY GTCGGAGCAA CCACCGCYTC GATACGCAGT 40
- (2) INFORMATION FOR SEQ ID NO:225:
- 10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 46 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:  
15 AGCCTTCACG TTCAGACCKC GTCGCCATCA AACRGTCAG ACCTGT 46
- (2) INFORMATION FOR SEQ ID NO:226:
- 20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 75 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:226:  
25 TCCCCCGCYG TGGGTATGGT GGTRGCGCAC RTYCTGCGDY TGCCCCAGAC 50  
CKTGTTYGAC ATAMTRGCGY GGGCC 75
- (2) INFORMATION FOR SEQ ID NO:227:
- 30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:227:  
ACGCCGGTGA CGCCTACAGT GGCTGTCGCA CACCCGGGC 39
- 35 (2) INFORMATION FOR SEQ ID NO:228:

- 229 -

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 42 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

ATGAGGGTCC CCACAGCCTT TCTCGACATG GTTGCCGGAG GC

42

- (2) INFORMATION FOR SEQ ID NO:229:

10

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

15

CGCGCCCTAT CCCAACGCAC CGTTAGAGTC CATGCGCAGG

40

- (2) INFORMATION FOR SEQ ID NO:230:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 49 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

20

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:230:

TCAGATCTTA CGGATCCCCT CTATCCTAGG TGA CTGCTC ACCGGGGGT

49

25

- (2) INFORMATION FOR SEQ ID NO:231:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 54 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:231:

CAGTCACGCT GCTGGGTGGC CCTTACTCCC ACCGTGGCGG YGYCTTATAT  
CGGT

50

54

35

- 230 -

## (2) INFORMATION FOR SEQ ID NO:232:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:232:

TAGCACTCTG GTRGAYCTAC TCRCTGGAGG G

31

## (2) INFORMATION FOR SEQ ID NO:233:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 54 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:233:

AAGTCTACAT GCTGGGTGTC TCTACCCCC ACCGTGGCTG CGCAACATCT  
GAAT

50

54

## (2) INFORMATION FOR SEQ ID NO:234:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

AGGCGCCATG GTCGACCTGC TTGCAGGCGG C

31

## (2) INFORMATION FOR SEQ ID NO:235:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

TCAGCCCCGA VYYTCGGAGC GGTCACGGCT CCTCTTCGGA GGG

43

- 231 -

## (2) INFORMATION FOR SEQ ID NO:236:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 44 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:236:

TGYTACGGAT YCCCCARGTG GTCATHGACA TCATWGCCGG GGSC

44

## (2) INFORMATION FOR SEQ ID NO:237:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

CATACCAAAT GCTTCCACGC CCGCAACGGG ATTCCGCAGG

40

## (2) INFORMATION FOR SEQ ID NO:238:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 37 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

TCTTCTTGCG GGCGCCGCAG TGGTTTGCTC ATCCCTG

37

## (2) INFORMATION FOR SEQ ID NO:239:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 52 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

ATCTAGCATC TTGAGGGTAC CTGAGATTG TGCGAGTGTG ATATTGGTG  
GC

50  
52

35

- 232 -

## (2) INFORMATION FOR SEQ ID NO:240:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

Trp	Ile	Gln	Val	Thr	Pro	Asn	Val	Ala	Val	Lys	His	Arg	Gly	Ala	
				5					10					15	
Leu	Thr	His	Asn	Leu	Arg	Xaa	His	Xaa	Asp	Xaa	Ile	Val	Met	Ala	
				20					25					30	
Ala	Thr	Val													

10

## (2) INFORMATION FOR SEQ ID NO:241:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

Trp	Val	Pro	Val	Ala	Pro	Asn	Leu	Ala	Ile	Ser	Gln	Pro	Gly	Ala	
				5					10					15	
Leu	Thr	Lys	Gly	Leu	Arg	Ala	His	Ile	Asp	Ile	Ile	Val	Met	Ser	
				20					25					30	
Ala	Thr	Val													

20

## (2) INFORMATION FOR SEQ ID NO:242:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:242:

Trp	Ile	Pro	Val	Xaa	Pro	Asn	Val	Ala	Val	Xaa	Xaa	Pro	Gly	Ala	
				5					10					15	
Leu	Thr	Gln	Gly	Leu	Arg	Thr	His	Ile	Asp	Met	Val	Val	Met	Ser	
				20					25					30	
Ala	Thr	Leu													

30

## (2) INFORMATION FOR SEQ ID NO:243:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids

35

- 233 -

(B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:243:

5 Trp Thr Xaa Val Thr Pro Thr Val Ala Val Arg Tyr Val Gly Ala  
                                   5                                  10                                  15  
 Thr Thr Ala Ser Ile Arg Ser His Val Asp Leu Leu Val Gly Ala  
                                   20                                  25                                  30  
 Ala Thr Xaa

10 (2) INFORMATION FOR SEQ ID NO:244:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:244:

Trp Val Ala Leu Xaa Pro Thr Leu Ala Ala Arg Asn Xaa Xaa Xaa  
                                   5                                  10                                  15  
 Xaa Thr Xaa Xaa Ile Arg Xaa His Val Asp Leu Leu Val Gly Ala  
                                   20                                  25                                  30  
 Ala Xaa Phe

20 (2) INFORMATION FOR SEQ ID NO:245:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

Trp Val Xaa Xaa Xaa Pro Thr Val Ala Thr Arg Asp Gly Lys Leu  
                                   5                                  10                                  15  
 Pro Xaa Xaa Gln Leu Arg Arg Xaa Ile Asp Leu Leu Val Gly Ser  
                                   20                                  25                                  30  
 30 Ala Thr Leu

(2) INFORMATION FOR SEQ ID NO:246:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

- 234 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

Trp Thr Pro Val Thr Pro Thr Val Ala Val Ala His Pro Gly Ala  
                             5                            10                            15  
 5 Pro Leu Glu Ser Phe Arg Arg His Val Asp Leu Met Val Gly Ala  
                             20                            25                            30  
 Ala Thr Leu

## (2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

Trp Val Ala Leu Thr Pro Thr Val Ala Xaa Xaa Tyr Ile Gly Ala  
                             5                            10                            15  
 15 Pro Leu Xaa Ser Xaa Arg Arg His Val Asp Leu Met Val Gly Ala  
                             20                            25                            30  
 Ala Thr Val

## (2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:  
 20 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

Trp Val Ser Leu Thr Pro Thr Val Ala Ala Gln His Leu Asn Ala  
                             5                            10                            15  
 25 Pro Leu Glu Ser Leu Arg Arg His Val Asp Leu Met Val Gly Gly  
                             20                            25                            30  
 Ala Thr Leu

## 30 (2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

35



- 235 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

Trp Val Pro Leu Thr Pro Thr Val Ala Ala Pro Tyr Pro Asn Ala  
5 10 15  
Pro Leu Glu Ser Met Arg Arg His Val Asp Leu Met Val Gly Ala  
20 25 30  
Ala Thr Met

## (2) INFORMATION FOR SEQ ID NO:250:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

Trp Val Xaa Ile Thr Pro Thr Leu Ser Ala Pro Xaa Xaa Gly Ala  
5 10 15  
Val Thr Ala Pro Leu Arg Arg Xaa Val Asp Tyr Leu Ala Gly Gly  
20 25 30  
Ala Ala Leu

## (2) INFORMATION FOR SEQ ID NO:251:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:251:

Trp His Ala Val Thr Pro Thr Leu Ala Ile Pro Asn Ala Ser Thr  
5 10 15  
Pro Ala Thr Gly Phe Arg Arg His Val Asp Leu Leu Ala Gly Ala  
20 25 30  
Ala Val Val

## (2) INFORMATION FOR SEQ ID NO:252:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

- 236 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:252:

Thr Leu Thr Met Ile Leu Ala Tyr Ala Ala Arg Val Pro Glu Leu  
                                   5                                  10                                  15  
 Xaa Leu Xaa Val Val Phe Gly Gly  
                                   20

5

## (2) INFORMATION FOR SEQ ID NO:253:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

Thr Thr Thr Met Leu Leu Ala Tyr Leu Val Arg Ile Pro Glu Val  
                                   5                                  10                                  15  
 Ile Leu Asp Ile Val Thr Gly Gly  
                                   20

15

## (2) INFORMATION FOR SEQ ID NO:254:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

Thr Xaa Thr Xaa Ile Leu Ala Tyr Xaa Met Arg Val Pro Glu Val  
                                   5                                  10                                  15  
 Ile Xaa Asp Ile Xaa Xaa Gly Ala  
                                   20

25

## (2) INFORMATION FOR SEQ ID NO:255:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:255:

Ala Val Gly Met Val Val Ala His Xaa Leu Arg Leu Pro Gln Thr  
                                   5                                  10                                  15  
 Xaa Phe Asp Ile Xaa Ala Gly Ala  
                                   20

35

- 237 -

## (2) INFORMATION FOR SEQ ID NO:256:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

Thr Xaa Ala Leu Val Xaa Ser Gln Leu Leu Arg Xaa Pro Gln Ala  
                    5                    10                    15  
Xaa Xaa Asp Xaa Val Xaa Gly Ala  
                    20

## (2) INFORMATION FOR SEQ ID NO:257:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

Thr Xaa Ala Leu Val Xaa Ala Gln Leu Leu Arg Xaa Pro Gln Ala  
                    5                    10                    15  
Xaa Leu Asp Met Ile Ala Gly Ala  
                    20

## (2) INFORMATION FOR SEQ ID NO:258:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:

Thr Thr Thr Leu Leu Leu Ala Gln Ile Met Arg Val Pro Thr Ala  
                    5                    10                    15  
Phe Leu Asp Met Val Ala Gly Gly  
                    20

## (2) INFORMATION FOR SEQ ID NO:259:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids  
(B) TYPE: amino acid

- 238 -

- (C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

5 Thr Thr Thr Leu Xaa Leu Ala Gln Val Met Arg Ile Pro Ser Thr  
5 10 15  
Leu Val Asp Leu Leu Xaa Gly Gly  
20

## (2) INFORMATION FOR SEQ ID NO:260:

- 10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

15 Thr Ala Thr Leu Val Leu Ala Gln Leu Met Arg Ile Pro Gly Ala  
5 10 15  
Met Val Asp Leu Leu Ala Gly Gly  
20

## (2) INFORMATION FOR SEQ ID NO:261:

- 20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:261:

25 Thr Ser Ala Leu Ile Met Ala Gln Ile Leu Arg Ile Pro Ser Ile  
5 10 15  
Leu Gly Asp Leu Leu Thr Gly Gly  
20

## (2) INFORMATION FOR SEQ ID NO:262:

- 30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:262:

35

- 239 -

Xaa Thr Ala Leu Xaa Met Ala Gln Xaa Leu Arg Ile Pro Gln Val  
5 10 15  
Val Ile Asp Ile Ile Ala Gly Xaa  
20

## (2) INFORMATION FOR SEQ ID NO:263:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:263:

10 Thr Thr Thr Leu Val Leu Ser Ser Ile Leu Arg Val Pro Glu Ile  
5 10 15  
Cys Ala Ser Val Ile Phe Gly Gly  
20

15

20

25

30

35

- 240 -

CLAIMS

1. A purified and isolated DNA having a sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:51.
2. A purified and isolated protein encoded by a gene whose sequence includes a sequence selected from the group consisting of SEQ ID NO:52 through SEQ ID NO:102.
3. A purified and isolated DNA having a sequence selected from the group consisting of SEQ ID NO: 103 through SEQ ID NO: 154.
4. A purified and isolated protein encoded by a gene sequence selected from the group consisting of SEQ ID NO: 155 through SEQ ID NO: 206.
5. A purified and isolated protein having an amino acid sequence selected from the group consisting of SEQ ID NO:52 through SEQ ID NO:102 and SEQ ID NO:155 through SEQ ID NO:206.
6. A method for the recombinant DNA-directed synthesis of a protein, said method comprising:
  - culturing a transformed or transfected host organism containing a DNA sequence capable of directing the host organism to produce said protein under conditions such that the protein is produced, said protein exhibiting substantial homology to a protein comprising the amino acid sequence selected from the group consisting of SEQ ID NO:52 through SEQ ID NO:102 or SEQ ID NO:155 through SEQ ID NO:206.
7. The method of claim 6, wherein the host

- 241 -

organism is transfected with a recombinant eukaryotic expression vector.

8. The method of claim 7, wherein the host organism is a eukaryotic cell.

9. A recombinant expression vector comprising a DNA sequence selected from the group consisting of SEQ ID NO:1 through SEQ ID NO:51 and SEQ ID NO:103 through SEQ ID NO:154.

10. A host organism transformed or transfected with a recombinant expression vector according to claim 9.

11. A method of detecting antibodies against HCV, said method comprising:

- (a) contacting a biological sample with at least one protein of claim 5 to form an immune complex with the antibodies; and
- (b) detecting the presence of the immune complex.

12. The method of claim 11 wherein the biological sample is selected from the group consisting of serum, saliva or lymphocytes or other mononuclear cells.

13. The method of claim 11, wherein the recombinant protein is bound to a solid support.

14. The method of claim 11, wherein the immune complex is detected using a labeled antibody.

15. A hepatitis C virus kit comprising: at least one protein comprising an amino acid sequence selected from the group consisting of: SEQ ID NO:52 through SEQ ID NO:102 and SEQ ID NO:155 through SEQ ID NO:206.

- 242 -

16. A composition comprising at least one recombinant protein of claim 5 and an excipient, diluent or carrier.

17. A composition comprising an expression vector capable of directing host organism synthesis of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO: 52 through SEQ ID NO: 102 and SEQ ID NO: 155 through SEQ ID NO: 206.

18. A method of preventing hepatitis C infection, comprising administering the composition of claim 16 or 17 to a mammal in an effective amount to stimulate the production of protective antibody.

19. A vaccine for immunizing a mammal against hepatitis C infection, comprising at least one protein according to claim 5 in a pharmacologically acceptable carrier.

20. A vaccine for immunizing a mammal against hepatitis C infection, said vaccine comprising an expression vector capable of directing host organism synthesis of a protein having an amino acid sequence selected from the group consisting of SEQ ID NO:52 - SEQ ID NO:102 and SEQ ID NO:155 - SEQ ID NO:206.

21. A method for detecting the presence of the hepatitis C virus via a reverse transcription-polymerase chain reaction, said method comprising amplifying an HCV reverse transcription product by polymerase chain reaction using universal primers.

22. The method of claim 21, wherein said universal primers are deduced from universally conserved nucleotide domains found in SEQ ID NO: 1 through SEQ ID NO:



- 243 -

51, in SEQ ID NO: 103 through SEQ ID NO: 154, or in consensus sequences shown in Figures 1A-H and 6A-K.

23. Substantially isolated and purified universal primers, wherein said primers have nucleic acid sequences derived from universally conserved nucleotide domains found in SEQ ID NO:1 through SEQ ID NO:51, in SEQ ID NO:103 through SEQ ID NO:154 and in consensus sequences showing Figures 1A-H and 6A-K.

24. A diagnostic kit for use in detecting the presence of hepatitis C virus in a biological sample, said kit comprising at least two universal primers according to claim 22.

25. A diagnostic kit for use in detecting the presence of hepatitis C virus in a biological sample, said kit comprising at least one nucleic acid sequence selected from the group consisting of SEQ ID No:1-51 or SEQ ID No:103-154.

26. A method for determining the genotype of a hepatitis C virus, said method comprising:

amplifying reverse transcription products of RNA via polymerase chain reaction using genotype-specific amplification primers deduced from genotype-specific nucleotide domains found in SEQ ID NO:1 through SEQ ID NO:51, in SEQ ID NO:103 through SEQ ID NO:154, or in consensus sequences shown in Figures 1A-H and 6A-K.

27. A method for determining the genotype of a hepatitis C virus, said method comprising:

(a) amplifying RNA via reverse

- 244 -

transcription-polymerase chain reaction  
to produce amplification products;

- (b) contacting said products with at least  
one sequence shown in SEQ ID NO:1  
through SEQ ID NO:51 and SEQ ID NO:103  
through SEQ ID NO:154; and
- (c) detecting complexes of said product  
which bind to said nucleic acid  
sequence.

10

28. A method for determining the genotype of a  
hepatitis C virus, said method comprising:

- (a) amplifying RNA via reverse  
transcription-polymerase chain reaction  
to produce amplification products;
- (b) contacting said products with at least  
one genotype-specific oligonucleotide;  
and
- (c) detecting complexes of said products  
which bind to said oligonucleotide(s).

15

20

29. The method of claims 27 or 28, wherein said  
amplification of step (a) uses universal primers deduced  
from universally conserved nucleotide domains found in SEQ  
ID NO:1 through SEQ ID NO:51, in SEQ ID NO:103 through SEQ  
ID NO:154, or in consensus sequences shown in Figures 1A-H  
and 6A-K.

25

30

30. The method of claim 28, wherein said  
genotype-specific oligonucleotide of step (b) is a nucleic  
acid sequence deduced from genotype-specific nucleotide  
domains found in SEQ ID NO:1 through SEQ ID NO:51 and SEQ  
ID NO:103 through SEQ ID NO:154, or in consensus sequences  
shown in Figures 1A-H and 6A-K.

35

31. Substantially isolated and purified

- 245 -

°  
genotype-specific oligonucleotides, wherein said  
oligonucleotides have nucleic acid sequences deduced from  
genotype-specific nucleotide domains found in SEQ ID NO:1  
through SEQ ID NO:51, in SEQ ID NO:103 through SEQ ID  
NO:154, or in consensus sequences shown in Figures 1A-H and  
5 6A-K.

32. Substantially purified and isolated  
genotype-specific peptides having amino acid sequences  
deduced from a genotype-specific amino acid domains located  
10 in SEQ ID NO:52 through SEQ ID NO:102, in SEQ ID NO:155  
through SEQ ID NO:206, or in consensus sequences shown in  
Figures 2A-H and 7A-K.

33. A method of detecting antibodies specific  
15 for a single genotype of HCV, said method comprising:  
(a) contacting a biological sample with at  
least one peptide of claim 32 to form  
an immune complex with the antibodies,  
and  
20 (b) detecting the presence of the immune  
complex.

34. The method of claim 33, wherein the  
biological sample is selected from the group consisting of  
25 serum, saliva or lymphocytes or other mononuclear cells.

35. The method of claim 33, wherein said peptide  
is bound to a solid support.

30 36. The method of claim 33, wherein the immune  
complex is detected using a labelled antibody or antigen.

37. A kit for use in detecting antibodies  
specific for a single genotype of HCV, said kit comprising:  
35 at least one peptide selected from the genotype-specific

- 246 -

peptides of claim 32.

38. Substantially purified and isolated universal peptides having amino acid sequences deduced from universally conserved amino acid domains found in SEQ ID NO:52 through SEQ ID NO:102, in SEQ ID NO:155 through SEQ ID NO:206, or in consensus sequences shown in Figures 2A-H and 7A-K.

39. A method of detecting antibodies against all genotypes of HCV, said method comprising:

- (a) contacting a biological sample with at least one peptide of claim 38 to form an immune complex with the antibodies, and
- (b) detecting the presence of the immune complex.

40. The method of claim 39, wherein the biological sample is selected from the group consisting of serum, saliva or lymphocytes or other mononuclear cells.

41. The method of claim 39, wherein said peptide is bound to a solid support.

42. The method of claim 39, wherein the immune complex is detected using a labelled antibody or antigen.

43. A composition comprising at least one peptide of claim 32 and an excipient, diluent or carrier.

44. A composition comprising at least one peptide of claim 38 and an excipient, diluent or carrier.

45. A method of preventing hepatitis C infection, comprising administering the composition of

- 247 -

° claims 43 or 44 to a mammal in an effective amount to stimulate production of a protective antibody.

46. A vaccine for immunizing a mammal against hepatitis C infection, comprising at least one peptide according to claims 32 or 38 in a pharmaceutically acceptable carrier.

47. A composition comprising at least one expression vector capable of directing host organism synthesis of a genotype-specific peptide having amino acid sequence deduced from a genotype-specific amino acid domain located in SEQ ID NO:52 - SEQ ID NO:102, and SEQ ID NO:155 - SEQ ID NO:206, or in consensus sequences shown in figures 2A-H and 7A-K.

48. A composition comprising at least one expression vector capable of directing host organism synthesis of a universal peptide having amino acid sequence deduced from universally conserved amino acid domains found in SEQ ID NO:52 - SEQ ID NO:102, and SEQ ID NO:155 - SEQ ID NO:206, or in consensus sequences shown in figures 2A-H and 7A-K.

49. A method of preventing hepatitis C infection, comprising administering the composition of claims 47 or 48 to a mammal in an effective amount to stimulate production of a protective antibody.

50. A vaccine for immunizing a mammal against hepatitis C infection, said vaccine comprising at least one expression vector capable of directing host organism synthesis of a geno-type specific peptide having amino acid sequence deduced from a geno type-specific amino acid domain located in SEQ ID NO:52 - SEQ ID NO:102, and SEQ ID NO:155 - SEQ ID NO:206, or in consensus sequences shown in

- 248 -

figures 2A-H and 7A-K.

51. A vaccine for immunizing a mammal against hepatitis C infection, comprising at least one expression vector capable of directing host organism synthesis of a universal peptide having amino acid sequence deduced from universally conserved amino acid domain found in SEQ ID NO:52 - SEQ ID NO:102, and SEQ ID NO:155 - SEQ ID NO:206, or in consensus sequences shown in figures 2A-H and 7A-K.

52. Anti-HCV core antibodies having specific binding affinity for core protein of a single genotype of HCV.

53. Anti-HCV envelope 1 antibodies having specific binding affinity for envelope 1 protein of a single genotype of HCV.

54. The antibodies of claims 52 or 53 wherein said antibodies are monoclonal antibodies.

55. A method of detecting core protein specific for a single genotype of HCV, said method comprising:

- (a) contacting a biological sample with at least one antibody of claim 52 to form an immune complex with said core protein, and
- (b) detecting the presence of the immune complex.

56. A method of detecting E1 protein specific for a single genotype of HCV, said method comprising:

- (a) contacting a biological sample with at least one antibody of claim 53 to form an immune complex with said E1 protein; and

- 249 -

- °
- (b) detecting the presence of the immune complex.

57. The methods of claims 55 or 56, wherein the biological sample is selected from the group consisting of serum, saliva lymphocytes or other mononuclear cells and liver.

58. The method of claims 55 or 56, wherein said antibody is bound to a solid support.

59. A method of detecting antibodies against all genotypes of HCV, said method comprising:

- (a) contacting a biological sample with at least one universal peptide of claim 38 to form an immune complex with said antibodies; and
- (b) detecting the presence of the immune complex.

1/89

FIGURE 1A

SEQ ID NO:	Isolate	
5	S14	1 TACCAAGTGC GCAACTCCACGGGGCTTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
1	DK7	1 TACCAAGTGC GCAACTCCACGGGGCTTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
8	US11	1 TACCAAGTGC GCAACTCCACGGGGCTTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
4	DR4	1 CACCAAGTGC GCAACTCTACAGGGCTTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
3	DR1	1 CACCAAGTGC GCAACTCTACAGGGCTTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
2	DK9	1 TACCAAGTACGCAACTCCTCGGGCCTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
6	S18	1 TACCAAGTACGCAACTCCTCGGGCCTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
7	SW1	1 TACCAAGTACGCAACTCCTCGGGCCTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
1-8	consensus	TACCAAGT- CGCAACTCcaCgGGGCTTACCATGTCACCAATGATTGCCCTAACTCGAGTA
5	S14	62 TtGTGTACGAGaCaGCTGATGCTATCCTaCACgCTCCGGGaTGTGTCCCTTGCGTTTCGtGA
1	DK7	62 TcGTGTACGAGGCGGCCGATGCCATCCTGCACACTCCGGGGTGTGTCCCTTGCGTTTCGCGA
8	US11	62 TTGTGTACGAGGCGGCCGATGCCATCCTGCACACTCCGGGGTGTGTTCCTTGCGTTTCGCGA
4	DR4	62 TTGTGTACGAGGCGGCCGATGCCATCCTGCACAGCCGGGGTGTGTTCCTTGCGTTTCGCGA
3	DR1	62 TTGTGTACGAGGCGGCCGATGCCATCCTGCACgCGCCGGGGTGTGTTCCTTGCGTTTCGCGA
2	DK9	62 TTGTGTACGAGGCGGCCGATGCCATCCTGCATCTCTCaGGGTGTGTTCCTTGCGTTTCGCGA
6	S18	62 TTGTGTACGAGACGGCCGATaCCATCCTACACTCTCCgGGGTGTGTTCCTTGCGTTTCGCGA
7	SW1	62 TTGTGTACGAGACGGCCGATgCCATCTACACTCTCCaGGGTGTGTTCCTTGCGTTTCGCGA
1-8	consensus	TtGTGTACGAGGcGcCGATgCaATcCTgCaC- CtCCgGGgTGTGTcCCTTGCGTTTCGcGA
5	S14	123 GGGTAACaCCTCGAGGTGTTGGGTGGCGATGACCCCCACGGTGGCCACCAGGGACGGCAAA
1	DK7	123 GGGTAACGtCTCGAGGTGTTGGGTGGCGATGACCCCCACGGTGGCCACCAGGGATGGCAAA
8	US11	123 GGGTAACGtCTCGAGGTGTTGGGTGGCGATGACCCCCACGGTGGCCACCAGGGACGGCAAA
4	DR4	123 GGGTAACaCCTCGAGGTGTTGGGTGGCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAA
3	DR1	123 GGGTAACGCTCTCGAGGTGTTGGGTGGCGGTGACCCCCACGGTGGCCACCAGGGACGGCAAA
2	DK9	123 GGGTAACGCTCTGAAATGTTGGGTGGCGGTGGCCCCACGGTGGCCACCAGGGACGGCAAg
6	S18	123 GGGTAACGCTCTGAgATGTTGGGTGGCGGTGGCCCCACAGTtGCCACCAGGGACGGCAAA
7	SW1	23 GGA TggCGCCcCGAaGTTGGGTGGCGGTGGCCCCACAGTcGCCACtAGGGACGGCAAA
1-8	consensus	GGgTaaCgcctCGAggTGTGGGTGGCGgTGaCCCCCAGgTgGCCACcAGGGACGGCAAA



2/89

FIGURE 1A

SEQ ID NO:	Isolate	
5	S14	184 CTCCCCgCAaCGCAGCTTCGACGTcACATCGATCTGCTtGTCGGGAGcGCCACCCTCTGTT
1	DK7	184 CTCCCCACA <sub>g</sub> CGCAGCTTCGACGTcACATCGATCTGCTcGTCGGGAGtGCCACCCTCTGTT
8	US11	184 CTCCCCACAACGCAaCTTCGACGTcACATCGATCTGCTTGTcGGGAGCGCCACCCTCTGTT
4	DR4	184 CTCCCCACAACGCAgCTTCGACGTcACATCGACCTGCTTGTcGGGAGCGCCACCCTCTGCT
3	DR1	184 CTCCCCACAACGCAgCTTCGACGTcACATCGACCTGCTTGTcGGGAGCGCCACCCTCTGCT
2	DK9	184 CTCCCCGCAACGCAgCTTCGACGTcACATCGATCTGCTTGTcGGGAGCGCCACCCTCTGCT
6	S18	184 CTCCCCGCAACGCAgCTTCGACGTcACATCGATCTGCTTGTtGGGAGCGCCACCCTCTGCT
7	SW1	184 CTCCctGCAACGCAgCTTCGACGTcACATCGATCTGCTTGTcGGaAGCGCCACCCTCTGCT
1-8	consensus	CTCCCC - CAaCGCAgCTTCGACGTcACATCGAtCTGCTtGTcGGgAGcGCCACCCTCTGct
5	S14	245 CGGCCCTCTACGTGGGGGACtTGTGCGGGTCTGTCTTTCTTGTcGGTCAgCTGTTTACCTT
1	DK7	245 CGGCCCTCTACGTGGGGGACCTGTGCGGGTCTGTCTTTCTTGTcGGTCAACTGTTTACCTT
8	S11	245 CGGCCCTCTACGTGGGGGACCTGTGCGGGTCTGTCTTTCTTGTcGGTCAACTGTTTACCTT
4	DR4	245 CGGCCCTCTACGTGGGGGACtTGTGCGGGTCTGTCTTCTTGTcGGTCAACTGTTTACCTT
3	DR1	245 CGGCCCTCTACGTGGGGGACcTGTGCGGGTCTGTCTTCTTGTcGGTCAACTGTTTACCTT
2	DK9	245 CGGCCCTCTATGTGGGGGACtTGTGCGGGTCTGTCTTCTTGTcGGCCAACTGTTTACCTT
6	S18	245 CGGCCCTCTATGTGGGGGACcTGTGCGGGTCTGTCTTTCTTGTcAGCCAgCTGTTTACctAT
7	SW1	245 CGGCCCTCTAcGTGGGGGACtTGTGCGGGTCTGTCTTTCTcGTcAGtCAaCTGTTTACgtT
1-8	consensus	CGGCCCTCTAcGTGGGGGAC - TGTGCGGGTCTGTCTTtCTtGTcGgGtCAaCTGTTtACctT
5	S14	306 CTCTCCcAGGCGCctCTGGACGACGCAAGaCTGCAATTGTTCTATCTATCCcGGCCATATA
1	DK7	306 CTCTCCcAGGCGCCACTGGACGACGCAAGGCTGCAATTGTTCTATCTATCCtGGCCATATA
8	S11	306 CTCTCCcAGaCGCCACTGGACGACGCAgGGCTGCAATTGTTCTATCTATCCCGGCCATATA
4	DR4	306 CTCTCCcAGGCaCCACTGGACAACGCAAGACTGCAATTGTTCTATCTATCCCGGCCATATA
3	DR1	306 tTCTCCcAGGCGCCACTGGACAACGCAAGACTGCAATTGTTCTATCTATCCCGGCCATATA
2	DK9	306 CTCCCCcAGaCGCCACTGGACAACGCAAGACTGCAACTGTTCTATCTATCCCGGCCATAtc
6	S18	306 CTCCCCcAGGCGCCACTGGACAACGCAAGACTGCAACTGTTCTATCTATCCCGGCCATATA
7	SW1	306 CTCCCCcAGGCGCCACTGGACAACGCAAGACTGtAACTGTTCTATCTAtCCCGGCCAcATA
1-8	consensus	cTCTcCCCAGgCgCCaCTGGACaACGCAaGACTGcAAtTGTTCTATCTAtCCcGGCCAtATA

3/89

FIGURE 1A

SEQ ID NO:	Isolate	
5	S14	367 ACGGGTCatCGCATGGCaTGGGATATGATGATGAACTGGTCCCCCTACgACGGCaCTGGTAG
1	DK7	367 ACGGGTCACCGCATGGCgTGGGATATGATGATGAACTGGTCCCCCTACcACGGCGTTGGTAG
8	S11	367 ACGGGTCACCGCATGGCaTGGGATATGATGATGAACTGGTCCCCCTACgCGGCGTTGGTgG
4	DR4	367 ACGGGcCACCGCATGGCgTGGGATATGATGATGAACTGGTCCCCCTACGACAGCGCTGGTAG
3	DR1	367 ACGGGaCACCGtATGGCaTGGGATATGATGATGAACTGGTCCCCCTACGACAGCGCTGGTAA
2	DK9	367 ACGGGTCatCGcatGGCgTGGGATATGATGATGAACTGGTCCCCCTACgCAGCGCTGGTAA
6	S18	367 ACGGGTCACCGtATGGCATGGGATATGATGATGAACTGGTCCCCCTACAACgCGcTGGTAA
7	SW1	367 ACGGGTCACCGcATGGCATGGGATATGATGATGAACTGGTCCCCcACAACaGCGcTGGTAg
1-8	consensus	ACGGGtCACCGcATGGCaTGGGATATGATGATGAACTGGTCCCCcACgaC-GCgcTGGTAg
5	S14	428 TAGCTCAGCTGCTCCGGATCCCcCAAGCCATCTTGGAtATGATCGCTGGTGCTCACTGGGG
1	DK7	428 TAGCTCAGCTGCTCCGGATCCCgCAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGG
8	S11	428 TAGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCTCACTGGGG
4	DR4	428 TAGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCCCACTGGGG
3	DR1	428 TGGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGaGCCCACTGGGG
2	DK9	428 TGGCgCAGCTGCTCAGGATCCCGCagGCCATCTTGGACATGATCGCTGGTGCCCACTGGGG
6	S18	428 TAGCTCAGCTGCTCAGGgTCCCGCAAGCCGCTTGGACATGATCGCTGGTGCCCACTGGGG
7	SW1	428 TAGCTCAGCTGCTCAGGaTCCCGCAAGCCGCTTGGACATGATCGCTGGTGCCCACTGGGG
1-8	consensus	TaGctCAGCTGCTcGGaTCCC-CAaGCCaTCTTGGAcATGATCGCTGGtGCCCACTGGGG
5	S14	489 AGTCCTaGCGGGCATAGCGTATTTcTCCATGGTGGGgAACTGGGCGAAGGTCTaGTgGTG
1	DK7	489 AGTCCTgGCGGGCATAGCGTATTTcTCCATGGTGGGGAACTGGGCGAAGGTCTGGTAGTG
8	S11	489 AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCTGGTAGTG
4	DR4	489 AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCTGGTAGTG
3	DR1	489 AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCTGGTAGTG
2	DK9	489 AGTCCTAGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCTGGTgGTa
6	S18	489 AGTCCTAGCGGGCATAGCGTATTTCTCCATGGcGGGGAACTGGGCGAAGGTCTGTcTAGTG
7	SW1	489 AGTCCTAGCGGGCATAGCGTATTTCTCCATGGtGGGGAACTGGGCGAAGGTCTGaTAGTG
1-8	consensus	AGTCCTaGCGGGCATAGCGTATTTcTCCATGGtGGGgAACTGGGCGAAGGTCTggtAGTg

4/89

FIGURE 1A

<u>SEQ ID NO:</u>	<u>Isolate</u>	
5	S14	550 CTGCTGCTATTcGCCGGCGTcGACGCG 
1	DK7	550 CTGCTGCTATTgCCGGCGTcGACGCG 
8	US11	550 CTGCTGCTATTgCCGGCGTcGACGCG 
4	DR4	550 CTGTTGCTGTTTGCCGGCGTTGATGCG 
3	DR1	550 CTGTTGCTGTTTGCCGGCGTTGATGCG 
2	DK9	550 CTGTTGCTGTTTaCCGGCGTcGATGCG 
6	S18	550 CTGTTGCTGTTTgCCGGCGTcGATGCG 
7	SW1	550 CTGTTGCTGTTTcCCGGCGTcGATGCG 
1-8	consensus	CTGcTGCTgTTtgCCGGCGTcGAtGCG

5/89

FIGURE 1B

<u>SEQ ID NO:</u>	<u>Isolate</u>	
11	DK1	1 TATGAAGTGCgCAACGTGTCCGGGgTGTACCACGTCACaAACGACTGCTCCAACCTCAAGCA
24	T10	1 TATGAAGTGCgCAACGTGTCCGGGaTGTACCAtGTCACgAACGACTGCTCCAACCTCAAGCA
10	D3	1 TATGAAGTGCgCAACGTGTCCGGGGTGTACCAaGTCACcAAcGACTGTTCCAACCTCGAGCA
9	D1	1 TATGAAGTGCgCAACGTGTCCGGGGTGTACCATGTTCACGAACGACTGTTCCAACCTCGAGCA
14	HK5	1 TATGAAGTGCgCAACGTGTCCGGGGTATACCATGTTCACGAACGACTGCTCCAACCTCAAGCA
15	HK8	1 TATGAAGTGCgCAACGTGTCCGGGATATACCATGTTCACGAACGACTGCTCCAACCTCAAGCA
12	HK3	1 TATGAAGTGCgCAACGTGTCCGGGATATACCATGTTCACGAACGACTGCTCCAACCTCAAGCg
23	T3	1 TAcGAAGTGCgCAACGTGTCCGGGGTGTACtATGTTCACGAACGACTGTTCCAACCTCAAGCA
22	SW2	1 TATGAAGTGCgCAACGTGTCCGGGGTGTAtCATGTTCACGAACGACTGTTCCAACCTCAAGCA
17	IND8	1 TATGAgGTGCgCAACGTGTCCGGGGTGTACCATGTTCACGAACGACTGCTCCAACCTCAAGTA
16	IND5	1 TATGAAGTGCgCAACGTGTCCGGGGTGTACCATGTTCACGAACGACTGCTCCAACCTCAAGTA
21	SA10	1 TATGAAGTGCgCAACGTGTCCGGGaTGTACCATGTTCACGAACGACTGCTCCAACCTCAAGCA
20	S45	1 TATGAAGTGCgCAACGTGTCCGGGgCGTACCATGTTCACGAACGACTGCTCCAACCTCAAGCA
25	US6	1 TATGAAGTGCgCAACGTGTCCGGGATGTACCATGTTCACGAACGACTGCTCCAACCTCAAGCA
13	HK4	1 cATGAAGTGCaCAACGTaTCCGGGATcTACCATGTTCACGAACGACTGCTCCAACCTCAAGTA
18	P10	1 TATGAAGTGCgCAACGTgTCCGGGGTGTACCATGTTCACGAACGACTGCTCCAACCTCAAGTA
19	S9	1 TATGAAGTGCgCAACGTaTCCGGGGcGTACCATGTTCACGAACGACTGCTCCAACCTCAAGTA
9-25	consensus	tAtGAaGTGCgCAACGTgTCCGGGgtgTAccAtGTCACgAACGACTGcTCCAACCTcaAGca

6/89

FIGURE 1B

<u>SEQ ID NO:</u>	<u>Isolate</u>	
11	DK1	62 TcGTGTaTGAGGCAGtGGACgTGATCATGCaTACCCCaGGGTGCGTGCCCTGCGTTCGGGA
24	T10	62 TtGTGTtTGAGGCAGCGGACtTGATCATGCACACCCCGGGTGCGTGCCCTGCGTTCGGGA
10	D3	62 TcGTGTATGAGACAGCGGACATGATCATGCACACCCCGGGTGCGTGCCCTGCGTTCGGGA
9	D1	62 TtGTGTATGAGACAGCGGACATGATCATGCACACCCCGGGTGCGTGCCCTGCGTTCGGGA
14	HK5	62 TCGTGTAcGAGACaAGGACATGATCATGCACACCCCTGGGTGCGTGCCCTGCGTTCGGGA
15	HK8	62 TCGTGTATGaaACAGCGGACATGATtATGCATACCCCTGGATGCaTGCCCTGCGTTCGGGA
12	HK3	62 TCGTGTATGAGACAGCaGACATGATCATGCATACCCCTGGATGCGTGCCCTGCGTaCGGGA
23	T3	62 TTGTGTATGAGACAGCGGACATGATCATGCACACCCCTGGGTGCGTGCCCTGCGTTCGGGA
22	SW2	62 TTGTGTATGAGACAGCGGACATGATCATGCAtACCCCGGGTGCGTGCCCTGCGTTCGGGA
17	IND8	62 TTGTGTATGAGGCAGCGGACATGATCATGCACACCCCGGGTGCGTGCCCTGCGTTCGGGA
16	IND5	62 TTGTGTATGAGGCAGCGGACATGATCATGCACACtCCCGGGTGCGTGCCCTGCGTTCGGGA
21	SA10	62 TTGTGTATGAGGCAGCGGACATGATCATGCACACCCCGGGTGCGTGCCCTGCGTTCGGGA
20	S45	62 TTGTGTATGAGGCAGtGGACgTGATCtTGACACCCCTGGGTGCGTGCCCTGCGTTCGGGA
25	US6	62 TTGTGTATGAGGCAGCGGACATGATCATGCACACtCCCGGGTGCGTGCCCTGtGTTTCGGGA
13	HK4	62 TTGTGTATGAGGCAGCGGACATGATCATGCAtACCCCGGGTGCGTGCCCTGtGTcCGGGA
18	P10	62 TTGTGTATGAGGCAGCGGACATGATaATGCACACCCCGGGTGCGTGCCCTGtGTTTCGGGA
19	S9	62 TTGTGTAcGAGGCAGCGGACgTGATcATGCAtACCCCGGGTGtGTaCCCTGcGTTCaGGA
9-25	consensus	TtGTGTatGAggCAGcgGACaTGATcaTGCACACcCCcGGgTGcgTgCCCTGcGTtCgGGA

7/89

FIGURE 1B

SEQ ID NO:	Isolate	
11	DK1	123 GaaCAACcaCTCCCGtTGCTGGGTAGCGCTCACeCCCACGCTCGCGGCCAGGAACgCCAGC
24	T10	123 GGgCAACTCCTCCCGCTGCTGGGTAGCGCTCACTeCCCACGCTCGCGGCCAGGAACACCAGC
10	D3	123 GGACAACTCCTCTCGCTGCTGGGTAGCGCTCACCCCCACGCTCGCGGCTAGGAATAGCAGC
9	D1	123 GGACAACTCCTCTCGCTGCTGGGTAGCGCTCACCCCCACGCTCGCGGCTAGGAATGGCAaC
14	HK5	123 aaACAACTCCTCCCGTTGtTGGGTAGCGCTCgCCCCACGCTCGCGGCCaAGGAACgCcCAGC
15	HK8	123 GAACAACTCCTCCCGTTGtTGGGTgGCGCTCACTCCCACGCTCGCGGCTAGGAATGTCAGC
12	HK3	123 GAACAACTCCTCCCGCTGtTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACGTCAGC
23	T3	123 GAgCAAtTCCTCCCGCTGCTGGGTAGCGCTLACTCCCACGCTCGCGGCCAGGAACGCCAGC
22	SW2	123 GGcCAACTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTaGCaGCCAGGAACaCCAGC
17	IND8	123 GGGCAACTtCTCTAgTtTGCTGGGTAGCGCTCACTCCCACTCTCGCGGCTAGGAACGCCAGC
16	IND5	123 GGGCAACTCCTCTCGCTGCTGGGTAGCGCTCACTCCCACTCTCGCGGCCAGGAACGCCAGC
21	SA10	123 GAACAACTCCTCCCGCTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACCTCAGC
20	S45	123 GAACAACTCCTCCCGtTGCTGGGTgGCGCTCACTCCCACGCTCGCGGCCAGGAACCTCAGC
25	US6	123 GAACAAAtTCCTCCCGeTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACGCtAGC
13	HK4	123 GAACAACTCCTCCCGtTGCTGGGTAGCGCTCACTCCCACGCTCGCGGCCAGGAACGCCAGC
18	P10	123 GAACAACTCCTCCCGeTGCTGGGTAGCGCTCACTCCCACaCTCGCGGCTAGGAAttCCAGC
19	S9	123 GggtAACTCCTCCCaTGCTGGGTgGCGCTCACeCCCACgCTCGCGGCCaAGGAACgCtACC
9-25	consensus	gaacAActcCTCccgTGCTGGGTAGCGCTCaCtCCCACgCTcGCgGCcAGGAACgCcAgC

8/89

FIGURE 1B

SEQ ID NO:	Isolate	
11	DK1	184 aTCCCCACTACGACaATACGACGCCATGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
24	T10	184 GTCCCCACTACGACgATACGACGCCATGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
10	D3	184 GTCCCCACTACGACaATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
9	D1	184 GTCCCCACTACGCGaATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
14	HK5	184 GTCCCCACcACGGCAATACGACGCCACGTCGACTTGCTCGTTGGGGCGGCTGCTTTCTGCT
15	HK8	184 GTCCCCACTACGACAATACGACGCCACGTCGACTTGCTCGTTGGGGCGGCTGCTTTCTGCT
12	HK3	184 GTCCCCACcACGACAATACGACGTCACGTCGACTTGCTCGTTGGGGCGGCTGCTTTCTGCT
23	T3	184 GTCCCCACTAaGACAATACGACGTCACGTCGACTTGCTCGTTGGGGCGGCTGCTTTCTGCT
22	SW2	184 GTCCCCACTACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
17	IND8	184 GTCCCCACCACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
16	IND5	184 GTCtCCACCACGACAATACGACaCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
21	SA10	184 GTCCCCACTACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
20	S45	184 GTCCCCACTACGACAATACGACGtCACGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
25	US6	184 GTCCCCACTACGACAATACGACGCCACGTCGATTTGCTCGTTGGGGCGGCTaCTTTCTGCT
13	HK4	184 aTCCCCACTACGACAATACGACGCCATGTCGAcTTGCTCGTTGGGGCGGCTGCTTTCTGCT
18	P10	184 GTCCCaACTACGgCAATACGACGCCATGTCGATTTGCTCGTTGGGGCGGCTGCTTTCTGCT
19	S9	184 GTCCCCACcACGaCAATACGACGtCATGTCGATTTGCTCGTTGGGGCGGCTGtTTCTGCT
9-25	consensus	gTCCcCACTAcGaCaATACGACgcCAcGTCGAtTTGCTCGTTGGGGCGGCTgctTTCTGCT

9/89

FIGURE 1B

SEQ ID NO.	Isolate	
11	DK1	245 CCGCTATGTACGTGGGgGACCTCTGCGGATCcgTTTTCTCGTCTCTCAGCTGTTACCTT 
24	T10	245 CCGCTATGTATGTGGGgGACCTCTGCGGATCTGTTTTCTCGTCTCTCAGCTGTTACCTT 
10	D3	245 CCGCCATGTACGTGGGGGATCTtTGCGGATCTGTTTTCTCGTCTCCAGCTGTTACCTT 
9	D1	245 CCGCCATGTACGTGGGGGATCTcTGCGGATCTGTTTTCTCTcATCTCCAGCTGTTACCCcT 
14	HK5	245 CCGCTATGTACGTGGGGGATCTtTGCGGATCTGTTTTCTCTcCTCTCCAGCTGTTACCTT 
15	HK8	245 CCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCTCGTCTCCAGCTGTTACCTT 
12	HK3	245 CCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCTCTtGTCTCCAGCTGTTACCTT 
23	T3	245 CCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCTCGTCTCCAGCTGTTCACTTT 
22	SW2	245 CCGtTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCTCTCGTCTCCAGCTGTTCACTTT 
17	IND8	245 CCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTTTCTCTtGTCTCCAGCTGTTACCTT 
16	IND5	245 CCGCTATGTACGTGGGGGATCTaTGCGGATCTGTTTTCTCTcGTCTCCAGCTGTTACCTT 
21	SA10	245 CCGCcATGTACGTGGGGGAcCTCTGCGGATCTGTTTTCTTGTCTCCAGCTGTTACCTT 
20	S45	245 CCGCTATGTACGTGGGGGAtCTCTGCGGATCTGTTTTCTCTTGTtTCCAGCTGTTACCTT 
25	US6	245 CCGCTATGTACGTGGGGGAcCTCTGCGGgTCcGTTTTCTCTcATCTCCAGCTGTTACCTT 
13	HK4	245 CCGCcATGTACGTGGGgGATCTCTGCGGATCTGTcTTCCTCGTCTCCAGtTGTTACCTT 
18	P10	245 CCGCTATGTACGTGGGGGATCTCTGCGGATCTGTTcTTCCTCGTCTCCAGCTGTTACCTT 
19	S9	245 CCGCTATGTACGTGGGGGAcTgTGCGGATCTGTTtTCTCTcATCTCCAGCTGTTACCCaT 
9-25	consensus	CCGctATGTACGTGGGgGATCTcTGCGGATCTcGTTtTCTCTcgTcTCCAGcTGTTACctT 



10/89

FIGURE 1B

SEQ ID NO:	Isolate	
11	DK1	306 tTCaCCTCGCCGGCATGAGACagcaCAGGACTGCAACTGCTCAATCTATCCCGGCCacgTt
24	T10	306 CTCGCCTCGCCGGCATGAGACttTgCAGGACTGCAACTGCTCAATCTATCCCGGCCatcTG
10	D3	306 CTCGCCTCGCCGGCATGAGACaGTACAGGAaTGTAAGTCTCAATCTATCCCGGCCACGTG
9	D1	306 CTCGCCTCGCCGGCATGAGACGGTACAGGagTGTAAtTGCTCAATCTATCCCGGCCACGTG
14	HK5	306 CTCGCCTCGCCGACACGAGACGGTACAGGACTGCAACTGCTCAATCTATCCCGGCCACGTA
15	HK8	306 tTCGCCTCGCCGACACGAGACGGTACAGGACTGCAACTGCTCAATCTATCCCGGCCACGTA
12	HK3	306 CTCGCCTCGCCGACACGAGACAGTACAGGACTGCAACTGCTCAaTCTATCCCGGCCACGTA
23	T3	306 CTCGCCTCGCCGGCAtGAGACAGTACAGGACTGCAACTGCTCAATCTATCCCGGCCACGTA
22	SW2	306 tTCACCTCGCCGGCacGAGACAGTACAGGACTGCAACTGtTCCATCTATCCCGGCCACGTA
17	IND8	306 CTCACCGCGCCGGCATGAGACAGTACAGGACTGCAATTGCTCCATCTATCCCGGCCACGTA
16	IND5	306 CTCACCGCGCCGGCATGAGACAGTACAGGACTGCAATTGCTCCATCTATCCCGGCCACGTA
21	SA10	306 CTCGCCTCGCCGGtATGAGACAGTACAGGACTGCAATTGCTCAATCTATCCCGGCCgCGTA
20	S45	306 CTCGCCTCGTcGGCATGAGACAGTACAGGACTGCAAcTGTTCaATCTATCCCGGCCACGTA
25	US6	306 CTCGCCTCGTcAGCATGAGACAGTACAGGACTGCAATTGtTCAATCTATCCCGGCCACGTA
13	HK4	306 CTCGCCTCGCCGGCATGAGACgGTACAGGACTGCAATTGcTCAATCTATCCCGGCCACGTA
18	P10	306 CTCaCCTCGCCGGCAttgGACAGTACAGGACTGCAATTGtTCAATCTATCCtGGGCCACGTA
19	S9	306 CTCgCCcCGtCGGCATgaGACAGTACAGaACTGCAATTGcTCAATCTATCCcGGaCACGTg
9-25	consensus	cTCgCCtCGcCggcAtgaGACagtaCAGgActGcAAcTGcTCaaTCTATCCcGGcCacgTa

11/89

FIGURE 1B

<u>SEQ ID NO:</u>	<u>Isolate</u>	
11	DK1	367 TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCaCCTACAACAGCcCTAGTGc
24	T10	367 TCAGGTCACCGCATGGCTTGGGAcATGATGATGAACTGGTCGCCTACAACAGCtCTAGTGG
10	D3	367 ACAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCGCCTACAgCAGCCCTAGTGG
9	D1	367 ACAGGTCACCGcATGGCTTGGGATATGATGATGAACTGGTCACCTACAACAGCCtTAGTGG
14	HK5	367 ACAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAACAGCCCTAGTGG
15	HK8	367 TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCgCCcACAACAGCCCTAGTGG
12	HK3	367 TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCcCCtACAgCAGCCCTAGTGG
23	T3	367 aCAGGTCACCGcATGGCTTGGGATATGATGATGAACTGGTCgCCcACAaCgGCaCTAGTGG
22	SW2	367 TCAGGTCACCGCATGGCTTGGGAcATGATGATGAACTGGTCACCTACAGCaGCCCTgGTGG
17	IND8	367 TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAGCgGCCCTAGTGG
16	IND5	367 TCAGGTCACCGCATGGCcTGGGATATGATGATGAACTGGTCACCTACAGCAGCCCTAGTGG
21	SA10	367 ACAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAaCAGCtCTAGTAg
20	S45	367 ACAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCgCCTACAGCAGCCtTAGTGG
25	US6	367 TCAGGTCACCGCATGGCTTGGGATATGATGATGAAtTGGTCACCTACAGCAGCCCTAGTGG
13	HK4	367 TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCACCTACAGCAGCCCTAGTGG
18	P10	367 TCAGGTCACCGCATGGCTTGGGATATGATGATGAACTGGTCGCCcACAGCAGCCCTAGTGG
19	S9	367 aCAGGTCAtcGCATGGCcTGGGATATGATGATGAACTGGTCGCCtACAaCAGCCCTAGTGG
9-25	consensus	tCAGGTCACCGcATGGCcTGGGATATGATGATGAACtGGTCaCCtACAgCaGCccTaGTgg

12/89

FIGURE 1B

<u>SEQ ID NO:</u>	<u>Isolate</u>	
11	DK1	428 TaTCGCAGTTACTCCGaaTCCCAACAAGCTGTCgTGGACATGGTgGcGGGGCCCCACTGGGG
24	T10	428 TgTCGCAGTTACTCCGGATCCCAACAAGCTGTCaTGGACATGGTGaCaGGGGCCCCACTGGGG
10	D3	428 TATCGCAGTTACTCCGGATCCCAACAAGCTGTCgTGGACATGGTGGCGGGGGCCCCACTGGGG
9	D1	428 TATCGCAGTTACTCCGGATCCCAACAAGCTGTCaTGGACATGGTGGCGGGGGCCCCACTGGGG
14	HK5	428 TGTCGCAGTTACTCCGGATCCCGCAAGCTGTCGTGGACATGGTaGCGGGGGCCCCACTGGGG
15	HK8	428 TGTCGCAGTTACTCCGGATCCCGCAAGCTaTCGTGGACATGGTGGCGGGGGCCCCACTGGGG
12	HK3	428 TGTCGCaATTACTCCGGATCCCGCAAGCTGTCGTGGACATGGTGGCGGGGGCCCCACTGGGG
23	T3	428 TGTCGCAGTTgTCCGGATCCCAACAAGCTGTCGTGGACATGGTGGCGGGGGCCCCACTGGGG
22	SW2	428 TATCGCAGTTaTCCGGATCCCAACAAGCTGTCGTGGACATGGTaGCGGGGGCCCCACTGGGG
17	IND8	428 TATCGCAGTTGTCCTCGGATCCCAACAAGCTGTCGTGGATATGGTGGCGGGGGCCCCACTGGGG
16	IND5	428 TATCGCAGTTGTCCTCGGATCCCAACAAGCTGTCGTGGATATGGTGGCGGGGGCCCCACTGGGG
21	SA10	428 TATCGCAGTTACTCCGGATCCCAACAAGCTaTCGTGGACATGGTGGCGGGGGCCCCACTGGGG
20	S45	428 TATCGCAGTTACTCCGGATCCCAACAAGCTGTCGTGGACATGGTGGCGGGGGCCCCACTGGGG
25	US6	428 TATCGCAGTTACTCCGGATCCCAACAAGCTGTCATGGACATGGTGGCGGGGGCCCCACTGGGG
13	HK4	428 TATCGCAGTTACTCCGaCTCCCAACAAGCTGTCATGGACATGGTGGCGGgaGCCCCACTGGGG
18	P10	428 TgTCGCAGCTACTCCGGATCCCAACAAGCTaTCtTGGATgTGGTGGCGGGGGCCCCACTGGGG
19	S9	428 TaTCGCAGCTACTCCGGATCCCAACAAGCTgTCaTGGATaTGGTGGCGGGGGCCCCACTGGGG
9-25	consensus	TaTCGCAGtTaCTCCGgaTCCCaCAAGCTgTCgTGGAcATGGTggCgGGgGCCCACTGGGG

13/89

FIGURE 1B

SEQ ID NO:	Isolate	
11	DK1	489 AGTCCTGGCGGGCCTcGCCTACTACTCCATGGCGGGGAAGTGGGcCAAGGTTTTAATTGTG
24	T10	489 AGTCCTGGCGGGCCTcGCCTACTATTCCATGGCGGGGAAGTGGGCTAAGGTTTTAATTGTG
10	D3	489 GGTCTGGCGGGCCTCGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
9	D1	489 GGTCTGGCGGGCCTCGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
14	HK5	489 GGTCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
15	HK8	489 AGTCCTAGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
12	HK3	489 AGTCCTAGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
23	T3	489 AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
22	SW2	489 AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
17	IND8	489 AATCCTGGCGGGCCTTGCCTACTATTCCATGGTAGGGGAAGTGGGCTAAGGTTTTGATTGTG
16	IND5	489 AATCCTGGCGGGCCTTGCCTACTATTCCATGGTAGGGGAAGTGGGCTAAGGTTTTGATTGTG
21	SA10	489 AGTCCTAGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTTGATTGTG
20	S45	489 AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTCTGATTGTG
25	US6	489 AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTCTGATTGTG
13	HK4	489 AGTCCTAGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGcCAAGGTTTTGATTGTG
18	P10	489 AGTCCTGGCGGGCCTTGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGCTTGATTGTG
19	S9	489 AGTCCTGGCGGGCCTcGCCTACTATTCCATGGTGGGGAAGTGGGCTAAGGTTTGATTGTG
9-25	consensus	agTCCTgGCGGGCCTcGCcTACTAtTCCATGGtgGGgAACTGGGCTAAGGTTtTgATTGTg

14/89

FIGURE 1B

SEQ ID NO:	Isolate	
11	DK1	550 tTGCTACTCCTTGCCGGCGTTGATGGG 
24	T10	550 ATGCTACTCCTTGCCGGCGTTGATGGG 
10	D3	550 ATGCTACTCCTTGCTGGCGTcGACGGC 
9	D1	550 ATGCTACTCCTTGCTGGCGTTGACGGC 
14	HK5	550 ATGCTACTcTTTGCCGGCGTTGATGGG 
15	HK8	550 ATGCTACTgTTTGCCGGCGTTGATGGG 
12	HK3	550 ATGCTACTcTTTGCCGGCGTTGATGGG 
23	T3	550 cTGCTACTCCTTGCCGGCGTTGATGGG 
22	SW2	550 ATGCTACTCCTTGcTGGCGTTGACGGG 
17	IND8	550 ATGCTACTCCTTGCCGGCGTTGACGGG 
16	IND5	550 ATGCTACTCCTTGCCGGCGTTGACGGG 
21	SA10	550 ATGCTACTCCTTGCCGGCGTTGACGGG 
20	S45	550 ATGCTACTCCTTGCCGGCGTTGACGGG 
25	US6	550 tTGCTACTCCTTGCCGGCGTTGACGGG 
13	HK4	550 ATGCTACTCCTTGCCGGCGTTGACGGG 
18	P10	550 ATGCTACTCCTTGCCGGCGTTGACGGa 
19	S9	550 ATGCTACTcTTTGcTGGTGTGACGGg 
9-25	consensus	aTGCTACTcCTTGCCGGcGTTGACGGg

15/89

FIGURE 1C

SEQ ID NO:	Isolate	
26	T2	1 GCcCAAGTGAgGAACACCAGccgCgGtTACATGGTGACTAACGACTGTTCcAATGAgAGCA 
27	T4	1 GCaCAAGTGAAGAACACCACtAaCAGCTACATGGTGACcAACGACTGTTCcAATGACAGCA 
28	T9	1 GCCgAAGTGAAGAACACCAGTACCAGCTACATGGTGACaAATGACTGTTCcAACGACAGCA 
29	US10	1 GtCcAAGTGAAaAACACCAGTACCAGCTAtATGGTGACcAATGACTGcTCCAACGACAGCA 
26-29	consensus	GcccAAGTGAagAACACCAGtacCaGcTAcATGGTGACcAA-GACTGtTCcAA-GACAGCA
SEQ ID NO:	Isolate	
26	T2	62 TCACcTGGCAGCTCCAaGCCGCGGTtCTCCACGTCCCCGGGTGTaTCCCGTGtGAGAggct 
27	T4	62 TCACtTGGCAGCTCCAGGCCGCGGTCTCCACGTCCCCGGGTGTGTCCCGTGCGAGAAaAc 
28	T9	62 TCACcTGGCAACTCCAGGCCGCGGTCTCCACGTCCCCGGGTGcGTCCCGTGCGAGAgAGT 
29	US10	62 TCACtTGGCAACTtgAGGctGCGGTCTCCACGTtCCCCGGGTGtGTCCCGTGCGAGAAAGT 
26-29	consensus	TCAC-TGGCA-CTccAgGcCGGTcCTCCACGTcCCCCGGGTgtGTCCCGTGcGAGA-agt
SEQ ID NO:	Isolate	
26	T2	123 GGGAAATACATCcCGaTGCTGGATACCGGTcaCACCAAACGTGGCCGTGCGGCAGCCCCGGC 
27	T4	123 GGGAAATACATCtCGGTCTGGATACCGGTtTCACCAAACGTGGCCGTGCGGCAGCCCCGGC 
28	T9	123 tGGAAAcgCgTCgCGGTCTGGATACCGGTCTCgCCAAACGTaGctGTGCAGCGGCCTGGC 
29	US10	123 gGGAAAtaCaTctCGGTCTGGATACCGGTCTCaCCAAAtGTgGcCGTCAGCGGCCTGGC 
26-29	consensus	gGGAAAtaCaTctCGgTGCTGGATACCGGTctCaCCAAAcGTgGcCGTGC-GC-GCC-GGC
SEQ ID NO:	Isolate	
26	T2	184 GCtCTtACGCAGGGCTTGCGGACGCACATcGACATGGTTGTGATGTCCGCCACGCTCTGCT 
27	T4	184 GCCCTCACGCAGGGCTTGCGGACGCACATtGACATGGTTGTGATGTCCGCCACGCTCTGCT 
28	T9	184 GCCCTCACGCAGGGCTTGCGGACGCACATCGACATGGTTGTGATGTCCGCCACGCTCTGCT 
29	US10	184 GCCCTCACGCAGGGCTTGCGGACTCACATCGACATGGTcGTGATGTCCGCCACGCTCTGCT 
26-29	consensus	GCcCTcACGCAGGGCTTGCGGACgCACATcGACATGGTtGTGATGTCCGCCACGCTCTGCT
SEQ ID NO:	Isolate	
26	T2	245 CTGCcCTtTACGTGGGGACCTCTGCGGCGGGGTGATGCTCGCAGCCCAGATGTTcATTtGT 
27	T4	245 CTGCTCTtTACGTGGGGACCTCTGCGGCGGGGTGATGCTCGCAGCCCAGATGTTcATTcGT 
28	T9	245 CCGCTCTtTACGTGGGGAtCTCTGCGGCGGGGTaATGCTCGCcGctCAGATGTTcATTaT 
29	US10	245 CCGCTCTtTACGTGGGGGActTCTGCGGtGGGAtgATGCTCGCaGcCcAaATGTTcATTgt 
26-29	consensus	C-GCtCT-TACGTGGGGGAcTCTGCGGcGGGgtgATGCTCGCaGcCcAgATGTTcATTgt

16/89

FIGURE 1C

<u>SEQ ID NO:</u>	<u>Isolate</u>	
26	T2	306 CTCGCCGcACgCCTCTGGTTTGTGCAAGAATGCAATTGCTCcATCTACCCcGGtACCATC 
27	T4	306 CTCGCCGCAACatCACTGGTTTGTGCAAGAcTGCAATTGCTCcATCTACCCTGGcACCATC 
28	T9	306 CTCGCCGcAgCACCACCTGGTTTGTGCAGGAATGCAACTGCTCCATtTACCCTGGTACCATC 
29	US10	306 CTCGCCGcGcCACCACCTcGTTTGTGCAGGAATGCAACTGCTCCATcTACCCcGGTACCATC 
26-29	consensus	CTCGCCGC - aCacCACTgGTTTGTGCA - GAaTGCAA - TGCTCcATcTACCC - GGtACCATC
<u>SEQ ID NO:</u>	<u>Isolate</u>	
26	T2	367 ACTGGACACCGTATGGCATGGGAcATGATGATGAACTGGTCGCCcACgGCCACCATGATCC 
27	T4	367 ACTGGACACCGTATGGCATGGGAtATGATGATGAACTGGTCGCCcACgGCCACCATGATCC 
28	T9	367 ACTGGACACCGTATGGCATGGGACATGATGATGAACTGGTCGCCcACaACCACCATGATCt 
29	US10	367 ACcGGgCACCGTATGGCATGGGACATGATGATGAACTGGTCGCCcACggCCACtTGATCc 
26-29	consensus	ACtGGaCACCGTATGGCATGGGAcATGATGATGAACTGGTCGCCcAC - gCCACcaTGATCc
<u>SEQ ID NO:</u>	<u>Isolate</u>	
26	T2	428 TGGCGTACGCGATGCGCGTTCCCGAGGTCATCaTAGACATCaTcgCGGGGcCtCACTGGGG 
27	T4	428 TGGCGTACGCGATGCGCGTTCCCGAGGTCATCtTAGACATCgTLAGCGGGGcCaCACTGGGG 
28	T9	428 TGGCGTACGCGATGCGCGTTCCCGAGGTCATCATAGACATCATcAGCGGaGCtCACTGGGG 
29	US10	428 TGGCGTACGtGATGCGCGTTCCCGAGGTCATCATAGACATCATLAGCGGgGCgCATTGGGG 
26-29	consensus	TGGCGTACGcGATGCGCGTTCCCGAGGTCATCaTAGACATCaT - aGCGGgGCtCAcTGGGG
<u>SEQ ID NO:</u>	<u>Isolate</u>	
26	T2	489 CGTCATGTTtGGCTTGGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAgGTCaTTGTTCATC 
27	T4	489 CGTCATGTTcGGCTTGGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAaGTCGTTGTTCATC 
28	T9	489 CGTCATGTTTCGGCtTAGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAgGTCGTTGTTCATC 
29	US10	489 CGTCtTGTTcGGCtTAGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAaGTCGTTGTTCATC 
26-29	consensus	CGTCaTGTTcGGCtT - GCCTACTTCTCTATGCAGGGAGCGTGGGCGAA - GTCgTTGTTCATC
<u>SEQ ID NO:</u>	<u>Isolate</u>	
26	T2	550 CTctTGCTGGCtGCTGGGGTGGACGCG 
27	T4	550 CTtctTGCTGGCCGCTGGGGTGGACGCG 
28	T9	550 CTgtTGCTcaCCGCTGGcGTGGACGCG 
29	US10	550 CTtctTGCTagCCGCTGGgGTGGACGCG 
26-29	consensus	CTt - TGCTggCcGCTGGgGTGGACGCG

17/89

FIGURE 1D

<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	1 GTGGAAGTTAGaAACAcCAGTTtTAGCTACTACGCCACCAATGATTGCTCgAACAAACAGCA
30	DK8	1 GTGGAAGTCAGGAACATCAGTTTcAGCTACTACGCCACCAATGATTGCTCAAAACAACAGCA
32	SW3	1 GTGGAAGTCAGGAACATCAGTTCTAGCTACTAtGCCACCAATGATTGCTCAAAACagCAGCA
31	DK11	1 GTGGAAGTCAGGAACAcCAGTTCTAGtTACTAcGCCACCAATGATTGCTCAAACAaCAGCA
30-33	consensus	GTGGAAGTCAGgAACa-CAGTTtTAGcTACTAcGCCACCAATGATTGCTCaAACAAcAGCA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	62 TCACCTGGCAgCTCACCaACGCAGTTCTCCACCTTCCGGATGCGTCCCATGTGAGAATGA
30	DK8	62 TCACCTGGCAACTCACCgACGCAGTTCTCCACCTTCCGGATGCGTCCCATGTGAGAATGA
32	SW3	62 TCACCTGGCAACTCACCACGCAGTtCTCCACCTTCCGGATGCGTCCCaGTGTGAGAATGA
31	DK11	62 TCACCTGGCAACTCACCACGCAGTtCTCCACCTTCCGGATGCGTCCCaGTGTGAGAATGA
30-33	consensus	TCACCTGGCAaCTCACCaACGCAGTtCTCCACCTTCCGGATGCGTCCCaGTGTGAGAATGA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	123 CAATGGCACCTtTGGCTGTCTGGATACAAGTgACACCTAATGTGGCTGTGAAACACCGtGGC
30	DK8	123 CAATGGCACCTTGCCTGTCTGGATACAAGTGACACCTAATGTGGCTGTGAAACACCGCGGC
32	SW3	123 CAATGGCACCTTGCCTGTCTGGATACAAGTGACACCTAATGTGGCTGTGAAACACCGCGGC
31	DK11	123 CAATGGCACCTTGCCTGTCTGGATACAAGTGACACCTAATGTGGCTGTGAAACACCGCGGC
30-33	consensus	CAATGGCACCTtTGC-CTGCTGGATACAAGTgACACCTAATGTGGCTGTGAAACACCGcGGC
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	184 GCACTcACTCacAACCTGCGAACgCATgTCGACGTGATCGTAATGGCAGCTACGGTCTGCT
30	DK8	184 GCACTtACTCAtAACCTGCGAACACACGTGACGTGATCGTAATGGCAGCTACGGTCTGCT
32	SW3	184 GCgCTCACTCACAACCTGCGAGCACACGTGATATGATCGTAATGGCAGCTACGGTCTGCT
31	DK11	184 GCaCTCACTCACAACCTGCGAGCACAtaTgATATGATtGTAAATGGCAGCTACGGTCTGCT
30-33	consensus	GCaCTcACTCacAACCTGCGA-CaCA-gTcGA--TGATcGTAAATGGCAGCTACGGTCTGCT
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	245 CGGCCTTGATGTGGGgGACGTgTGCGGGGCCGTGATGATAgcGTGCGAGGCTtTCATAAT
30	DK8	245 CGGCCTTGATGTGGGAGACGTaTGCGGGGCCGTGATGATCGTGTGCGAGGCTtTCATAAT
32	SW3	245 CGGCCTTGATGTGGGAGACaTGTCGGGGCCGTGATGATCGTGTGCGAGGCTTTTCATAAT
31	DK11	245 CGGCCTTGATGTGGGAGACgTGTCGGGGCCGTGATGATCGTGTGCGAGGCTTTTCATAgT
30-33	consensus	CGGCCTTGATGTGGGgAGACgTgTGCGGGGCCGTGATGATcGtGTGCGAGGCTtTCATAaT



18/89

FIGURE 1D

<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	306 ATCGCCaGAACGCCACAACCTTcACCCAGGAGTGCAACTGTTCCATCTACCAAGGTCATATC
30	DK8	306 ATCGCCcGAACGCCACAACCTTTACCCAGGAGTGCAACTGTTCCATCTACCAAGGTCATATC
32	SW3	306 ATCGCCAGAAGGCCACAACCTTTACCCAGGAGTGCAACTGTTCCATCTACCAAGGTCgTATC
31	DK11	306 ATCGCCAGAACaCCACcACTTTACCCAGGAGTGCAACTGTTCCATCTACCAAGGTCaCATC
30-33	consensus	ATCGCCaGAACgCCACaACTTtACCCA - GAGTGCAACTGTTCCATCTACCAAGGTCatATC
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	367 ACCGGCCACCGCATGGCATGGGACATGATGCTgAACTGGTCACCAACTCTcACCATGATCC
30	DK8	367 ACCGGCCACCGCATGGCATGGGACATGATGCTAAACTGGTCACCAACTCTtACCATGATCC
32	SW3	367 ACCGGCCACCGCATGGCgTGGGACATGATGCTAAACTGGTCACCAACTCTtACCATGATCC
31	DK11	367 ACCGGCCACCGCATGGCaTGGGACATGATGCTtAACTGGTCACCAACTCTcACCATGATCC
30-33	consensus	ACCGGCCACCGCATGGCaTGGGACATGATGCTaAACTGGTCACCAACTCT - ACCATGATCC
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	428 TCGCCTAcGCtGCTCGTGTgCCTGAaCTAGtCCTtgaAGTTGTCTTCGGCGGCCATTGGGG
30	DK8	428 TCGCCTATGCCGCTCGTGTTCCTGAGCTAGcCCTccAgGTTGTCTTCGGCGGCCATTGGGG
32	SW3	428 TtGCCTATGCCGCTCGTGTTCCTGAGCTAGTCCTTGAAGTTGTCTTCGGCGGCCATTGGGG
31	DK11	428 TcGCCTATGCCGcCGTGTTCCTGAGCTAGTCCTTGAAGTcGTCTTCGGtGGtCATTTGGGG
30-33	consensus	TcGCCTATGCcGCtCGTGTtCCTGAgCTAGtCCTtgaAGTtGTCTTCGGcGGcCATTTGGGG
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	489 CGTGGTGTtTGGCTTGGCCTATTTCTCCATGCAaGGAGCGTGGGCCAAAGTCATcGCCATC
30	DK8	489 CGTGGTGTtTGGCTTGGCCTATTTCTCCATGCAgGGAGCGTGGGCCAAAGTCATTGCCATC
32	SW3	489 CGTGGTGTtTGGCTTGGCCTATTTCTCCATGCAaGGAGCGTGGGCCAAAGGTCATTGCCATC
31	DK11	489 tGTGGTGTtTGGCTTGGCCTATTTCTCCATGCAgGGAGCGTGGGCCAAGGTCATTGCCATC
30-33	consensus	cGTGGTGTtTGGCTTGGCCTATTTCTCCATGCA - GGAGCGTGGGCCAA - GTCATtGCCATC
<u>SEQ ID NO:</u>	<u>Isolate</u>	
33	T8	550 CTCCTcCTTGTcGCAGGAGTGGAcGCA
30	DK8	550 CTCCTtCTTGTcGCAGGAGTGGATGCA
32	SW3	550 CTCCTgCTTGTcGCAGGAGTGGATGCA
31	DK11	550 CTCCTtCTTGTaGCAGGAGTGGATGCA
30-33	consensus	CTCCTtCTTGTcGCAGGAGTGGAtGCA

19/89

FIGURE 1E

<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	1 tTAGAGTGGCGGAATGTGTCCGGCCTCTAcGTCCTTACCAACGACTGTtCCAATAGCAGTA
36	HK10	1 CTAGAGTGGCGGAATGTGTCTGGCCTCTATGTCTTACCAACGACTGTtCCAATAGCAGTA
37	S2	1 CTAGAGTGGCGGAATACGTCTGGCCTCTATGTCTTACCAACGACTGTtCCAATAGCAGTA
39	S54	1 CTAGAGTGGCGGAATACGTCTGGCCTCTATtTCCTTACCAACGACTGTtCCAATAGCAGTA
38	S52	1 CTAGAGTGGCGGAATACGTCTGGCCTCTATgTCCTTACCAACGACTGTtCCAATAGCAGTA
35-39	consensus	cTAGAGTGGCGGAATAcGTCtGGCCTCTAtgTCCTtACCAACGACTGTtCCAATAGCAGTA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	62 TcGTGTATGAGGCCGATGACGTcATTCTGCACACACCTGGCTGTGTACCTTGTGTTCAGGA
36	HK10	62 TTGTGTATGAGGCCGATGACGTcATTCTGCACACACCTGGCTGTGTACCTTGTGTTCAGGA
37	S2	62 TTGTGTATGAGGCCGATGACGTtATTCTGCACACACCTGGCTGTGTACCTTGTGTTCAGGA
39	S54	62 TTGTGTATGAGGCCGATGACGTcATTCTGCACACACCCGGCTGTGTACCTTGTGTTCAGGA
38	S52	62 TTGTGTATGAGGCCGATGACGTcATTCTGCACACACCCGGCTGTGTACCTTGTGTTCAGGA
35-39	consensus	TtGTGTATGAGGCCGATGACGTcATTCTGCACACACCTGGCTGTGTACCTTGTGTTCAGGA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	123 CGGCAATACATCtACGTGCTGGACCTCaGTGACgCCTACAGTGGCAGTCAGGTACGTTCGGA
36	HK10	123 CGGCAATACATCCACGTGCTGGACCTCgGTGACACCTACAGTGGCAGTCAGGTACGTTCGGA
37	S2	123 CGGtAATACATCCACGTGCTGGACCCcAGTGACACCTACAGTGGCAGTCAGGTAtGTTCGGA
39	S54	123 CGGCAATACATCCACGTGCTGGACCCcAGTGACACCTACGGTGGCAGTCAGGTACGTTCGGA
38	S52	123 CGGCAATACATCCAtGTGCTGGACCCcAGTGACACCTACGGTGGCAGTCAGGTACGTTCGGA
35-39	consensus	CGGcAATACATCcAcGTGCTGGACCCcCaGTGACaCCTACaGTGGCAGTCAGGTACGTTCGGA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	184 GCAACCACCGCtTCGATACGCAGTCATGTGGACCTGcTAGTGGGCGCGGCCACGATGTGCT
36	HK10	184 GCAACCACCGCtTCGATACGCAGTCATGTGGACCTGtTAGTGGGCGCGGCCACGATGTGCT
37	S2	184 GCAACCACCGCTTCGATACGCAGTCATGTGGACCTATTgGTGGGCGCGGCCACtATGTGCT
39	S54	184 GCAACCACCGCTTCGATACGCAGTCATGTGGACCTATTAGTGGGCGCGGCCACGCTGTGCT
38	S52	184 GCAACCACCGCTTCGATACGCAGTCATGTGGACCTATTAGTGGGCGCGGCCACGCTGTGCT
35-39	consensus	GCAACCACCGCtTCGATACGCAGTCATGTGGACCTAtTAgTGGGCGCGGCCACGaTGTGCT

20/89

FIGURE 1E

<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	245 CTGCGCTCTACGTGGGtGATgTGTGTGGGGCCGTCTTCCTtGTGGGACAAGCCTTCACGTT
36	HK10	245 CTGCGCTCTACGTGGGcGATATGTGTGGGGCCGTCTTCCTCGTGGGACAAGCCTTCACGTT
37	S2	245 CTGCGCTCTACGTGGGTGATATGTGTGGGGCCGTCTTCCTCGTGGGACAAGCCTTCACGTT
39	S54	245 CTGCGCTCTATGTGGGTGATATGTGTGGGGCCGTCTTCCTCGTGGGACAAGCCTTCACGTT
38	S52	245 CTGCGCTCTATGTGGGTGATATGTGTGGGGCCGTCTTCCTCGTGGGACAAGCCTTCACGTT
35-39	consensus	CTGCGCTCTAcGTGGGtGATaTGTGTGGGGCCGTCTTtCTcGTGGGACAAGCCTTCACGTT

<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	306 CAGACcTcGTcGCCATCAAACaGTCCAGACCTGTAACtGCTcGCTGTACCCAGGCCAtCTT
36	HK10	306 CAGACcGcGTcGCCATCAAACGGTCCAGACCTGTAACtGCTcGCTGTACCCAGGCCAcCTT
37	S2	306 CAGACCTCGTcGCCATCAAACGGTCCAGACCTGTAACtGCTcGCTGTACCCAGGCCATCTT
39	S54	306 CAGACCTCGTcGCCATCAAACGGTCCAGACCTGTAACtGCTcGCTGTACCCAGGCCATCTT
38	S52	306 CAGACCTCGTcGCCATCAAACGGTCCAGACCTGTAACtGCTcGCTGTACCCAGGCCATgTT
35-39	consensus	CAGACcTcGTcGCCATCAAACgGTCCAGACCTGTAACtGCTcGCTGTACCCAGGCCAtcTT

<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	367 TCAGGACATCGAATGGCTTGGGATATGATGATGAATTGGTCCCCCGctGTGGGTATGGTGG
36	HK10	367 TCAGGACATCGAATGGCTTGGGATATGATGATGAATTGGTCCCCCGcGTGGGTATGGTGG
37	S2	367 TCAGGACATCGcATGGCTTGGGATATGATGATGAATTGGTCCCCCGCTGTGGGTATGGTGG
39	S54	367 TCAGGACATCGAATGGCTTGGGATATGATGATGAATTGGTCCCCCGCTGTGGGTATGGTGG
38	S52	367 TCAGGACATCGAATGGCTTGGGATATGATGATGAATTGGTCCCCCGCTGTGGGTATGGTGG
35-39	consensus	TCAGGACATCGaATGGCTTGGGATATGATGATGAATTGGTCCCCCGctGTGGGTATGGTGG

<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	428 TaGCGCACGTcCTGCGtTtTGCCCCAGACCTTGTTCGACATAATAGctGGGGCCcATTGGGG
36	HK10	428 TGGCGCACGTcCTGCGgTTGCCCCAGACCTTGTTCGACATAATAGCCGGGGCCcATTGGGG
37	S2	428 TGGCGCACGTtCTGCGtTTGCCCCAGACcGtGTTCGACATAATAGCCGGGGCCcATTGGGG
39	S54	428 TGGCGCACATcCTGCGATTGCCCCAGACCTTGTTCGACATACTGGCCGGGGCCcATTGGGG
38	S52	428 TGGCGCACATcCTGCGATTGCCCCAGACCTTGTTCGACATACTGGCCGGGGCCcATTGGGG
35-39	consensus	TgGCGCACgtcCTGCG- tTGCCCCAGACcTGTTCGACATAaTaGcCGGGGCCcATTGGGG

21/89

FIGURE 1E

<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	489 CATCaTGGCgGGCCTAGCCTATTACTCCATGCAGGGCAACTGGGCCAAGGTCGCTATCATC 
36	HK10	489 CATCTTGGCaGGCCTAGCCTATTACTCCATGCAGGGCAACTGGGCCAAGGTCGCTATCATC 
37	S2	489 CATCTTGGCGGGCCTAGCCTATTACTCCATGCaaGGCAACTGGGCCAAGGTCGCTATCATC 
39	S54	489 CATCTTGGCGGGCCTAGCCTATTATTCTATGCAGGGCAACTGGGCCAAGGTCGCTATCATC 
38	S52	489 CATCTTGGCGGGCCTAGCCTATTATTCTATGCAGGGCAACTGGGCCAAGGTCGCTATtGTC 
35-39	consensus	CATCtTGGCgGGCCTAGCCTATTaTCcATGCagGGCAACTGGGCCAAGGTCGCTATcATC
<u>SEQ ID NO:</u>	<u>Isolate</u>	
35	DK12	550 ATGGTTATGTTTTTCAGGaGTCGATGCC 
36	HK10	550 ATGGTTATGTTTTTCAGGGGTCGATGCC 
37	S2	550 ATGGTTATGTTTTTCAGGGGTCGacGCC 
39	S54	550 ATGATTATGTTTTTCAGGGGTCGATGCC 
38	S52	550 ATGATTATGTTTTTCAGGGGTCGATGCC 
35-39	consensus	ATGgTTATGTTTTTCAGGgGTCGAtGCC

22/89

FIGURE 1F

<u>SEQ ID NO:</u>	<u>Isolate</u>	
43	27	1 GTCAACTATC <del>CAATGCCTCGGGCGTCTATCACATC</del> ACCAACGACTGCCCGAACTCGAGCA
42	26	1 GTCAACTATC <del>CGCAATGCCTCGGGCGTCTATCACGTC</del> ACCAACGACTGCCCGAACTCGAGCA
42-43 consensus (26)		GTCAACTATC <del>gCAATGCCTCGGGCGTCTATCACgTC</del> ACCAACGACTGCCCGAACTCGAGCA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
43	27	62 TAA <del>TGTATGAGGCCGAACACCACATCCTACACCTCCAGGGTGCGTAC</del> CCTGTGTGAGGGA
42	26	62 TAGTGTATGAGGCCGAACACCA <del>gATCTTACACCTCCAGGGTGCTg</del> CCCTGTGTGAGGGT
42-43 consensus (26)		TAgTGTATGAGGCCGAACACCA <del>gATCtTACACCTCCAGGGTGCTg</del> CCCTGTGTGAGGGT
<u>SEQ ID NO:</u>	<u>Isolate</u>	
43	27	123 gGGGAACCA <del>GTCACGCTGCTGGGTGGCCCTTACTCCACCGTGGCGGcGc</del> CTTATATCGGT
42	26	123 tGGGAAtCAGTCACGCTGCTGGGTGGCCCTTACTCCACCGTGGCGG <del>tg</del> cCTTATATCGGT
42-43 consensus (26)		tGGGAAtCAGTCACGCTGCTGGGTGGCCCTTACTCCACCGTGGCGG <del>tg</del> cCTTATATCGGT
<u>SEQ ID NO:</u>	<u>Isolate</u>	
43	27	184 GCaCCGCTTGA <del>aTCCaTCCGGAGACATGTGGACCTGATGGTAGGCGCt</del> GCTACaGTGTGCT
42	26	184 GCTCCGCTTGA <del>cTCCcTCCGGAGACATGTGGACCTGATGGTGGGCGCCGCTACTGTa</del> TGCT
42-43 consensus (26)		GCTCCGCTTGA <del>cTCCcTCCGGAGACATGTGGACCTGATGGTg</del> GGCGCcGCTACTGTaTGCT
<u>SEQ ID NO:</u>	<u>Isolate</u>	
43	27	245 CcGCTCTCTACa <del>TTGGGGACCTGTGCGGTGGcGtATTtTTGGTTGGc</del> CAGATGTTtTCTTT
42	26	245 CtGCCCTCTACg <del>TTGGAGAtCTGTGCGGTGGTgcATTCTTGGTTGGc</del> CAGATGTTCTCCTT
42-43 consensus (26)		CtGCCCTCTACg <del>TTGGaGAtCTGTGCGGTGGtGcATTCTTGGTTGGc</del> CAGATGTTtTCcTT
<u>SEQ ID NO:</u>	<u>Isolate</u>	
43	27	306 CCAGCCGCGACGCCACTGGACTACGCAGGACTGCAATTGTTCCATCTAtGCgGGGCagTt
42	26	306 CCAGCCGCGACGCCACTGGACTACGCAGGACTGCAATTGTTcAtCTACGCAGGGCATATC
42-43 consensus (26)		CCAGCCGCGACGCCACTGGACTACGCAGGACTGCAATTGTTcAtCTAcGCaGGGCAtaTc
<u>SEQ ID NO:</u>	<u>Isolate</u>	
43	27	367 ACaGGCCACAGa <del>ATGGCATGGGACATGATGATGAACTGGAGTCCCACAACCACc</del> TGgTCC
42	26	367 ACgGGCCACAGgATGGCATGGGACATGATGATGAACTGGAGTCCCACAACCACCTGcTtC
42-43 consensus (26)		ACgGGCCACAGgATGGCATGGGACATGATGATGAACTGGAGTCCCACAACCACcTGcTtC

23/89

FIGURE 1F

SEQ ID NO: Isolate

43 27

42 26

42-43 consensus (26)

428 TCGCCCAGGTcATGAGGATCCCTAGCACTCTGGTgGACCTACTCaCTGGAGGGCACTGGGG  
 |||||  
 428 TCGCCCAGGTcATGAGGATCCCTAGCACTCTGGTaGAtCTACTCGCTGGAGGGCACTGGGG  
 TCGCCCAGGTcATGAGGATCCCTAGCACTCTGGTaGAtCTACTCgCTGGAGGGCACTGGGG

SEQ ID NO: Isolate

43 27

42 26

42-43 consensus (26)

489 taTCCTTaTcGGGgTGGCaTACTTcGCATGCAAGCTAATTGGGCCAAGGTCAtcCTGGTC  
 |||||  
 489 CgTCCTTGTtGGGgTGGCGTACTTCAGtATGCAAGCTAATTGGGCCAAaGTCATcCTGGTC  
 cgTCCTTgTtGGGgTGGCGTACTTCAGtATGCAAGCTAATTGGGCCAAaGTCATcCTGGTC

SEQ ID NO: Isolate

43 27

42 26

42-43 consensus (26)

550 CTTTTCTCTcCGCTGGAGTTGATGCC  
 |||||  
 550 CTTTTCTCTcCGCTGGAGTTGATGCC  
 CTTTTCTCTcCGCTGGAGTTGATGCC

24/89

FIGURE 1G

<u>SEQ ID NO:</u>	<u>Isolate</u>	
45	SA1	1 GTtCCCTACCGgAATGCCTCTGGGGTTTAcCATGTCACCAATGAcTGCCCAAACTCtTCCA
47	SA5	1 GTCCCCCTACCGAAATGCCTCTGGGGTTTATCATGTCACCAATGATTGCCCAAACTCTTCCA
49	SA7	1 GTCCCCCTACCGAAATGCCTCgGGGGTTTATCATGTCACCAATGATTGCCCGAACTCTTCCA
46	SA4	1 GTTCCCTACCGAAAcGCCTCTGGGGTTTATCATGTCACCAATGATTGCCCAAACTCTTCCA
50	SA13	1 GTTCCCTACCGAAATGCCTCTGGGGTTTATCATGTCACCAATGATTGCCCAAACTCTTCCA
48	SA6	1 GTTCCcTACCGgAATGCCTCTGGGGTgTATCATGTcACCAATGATTGCCCAAACTCTTCCA
45-50	consensus	GTtCCcTACCGAAAtGCCTCtGGGGTtTAtCATGTcACCAATGAcTGCCCAAACTCtTCCA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
45	SA1	62 TAGTCTACGAGGCTGATAgCCTGATctTGCACGCACCTGGcTGCCTGCCCTGTGTCAgGcA
47	SA5	62 TAGTCTACGAGGCTGATAACCTGATcCTGCACGCACCTGGTTGCGTGCCCTGTGTCAaGgA
49	SA7	62 TAGTCTAtGAGGCTGAaACCTGATCCTGCACGCACCTGGTTGCGTGCCCTGTGTCAgAcA
46	SA4	62 TAGTtTACGAGGCTGATAACCTGATCTTGCAcGCACCTGGTTGCGTGCCcTGTGTcAGGCA
50	SA13	62 TcGTCTACGAGGCTGATGACCTGATCTTACACGCACCTGGTTGCGTGCCCTGTGTtAGGCA
48	SA6	62 TaGTCTAtGAGGCTGATGACCTGATCcTACACGCACCTGGcTGCCTGCCCTGTGTccGGaA
45-50	consensus	TaGTcTACGAGGCTGATaaCCTGATc-TgCAcGCACCTGGcTGCCTGCCcTGTGTcaggcA
<u>SEQ ID NO:</u>	<u>Isolate</u>	
45	SA1	123 AGaTAATGTCAGTAGGTGCTGGGTCCAAATCACCCCCACaTGTCAGCCCCGAcCtTCGGA
47	SA5	123 AGgTAATGTCAGTAGGTGCTGGGTCCAAATCACCCCCACATTGTcAGCCCCGAACCTCGGA
49	SA7	123 AaATAATGTCAGTAGGTGCTGGGTCCAAATCACCCCCACATTGTcAGCCCCGAACCTCGGA
46	SA4	123 AGATAATGTCAGTAaGTGCTGGGTCCAAATCACCCCCACgTTGTcAGCCCCGAAtCTCGGA
50	SA13	123 GGgTAATGTCAGTAGGTGCTGGGTCCAgATCACCCCCACACTGTcAGCCCCGAGCCTCGGA
48	SA6	123 GGaTAATGTCAGTAGaTGCTGGGTtCAcATCACCCCCACACTaTCAGCCCCGAGCCTCGGA
45-50	consensus	agaTAATGTCAGTAggTGCTGGGTcCAaATCACCCCCACa-TgTCAGCCCCGAaccTCGGA

25/89

FIGURE 1G

SEQ ID NO:	Isolate	
45	SA1	184 GCGGTCACGGCTCCTCTTCGGAGGGcCGTTGACTACTTAGCGGGAGGgGCTGCTcCTCTGCT
47	SA5	184 GCGGTCACGGCTCCTCTTCGGAGGGtCGTTGACTACTTAGCGGGAGGGGCTGCCCTCTGCT
49	SA7	184 GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGACTACcTAGCGGGAGGGGCTGCCCTCTGCT
46	SA4	184 GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGACTACTTAGCGGGAGGGGCTGCCCTCTGCT
50	SA13	184 GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGACTACTTAGCGGGgGGGGCTGCCCTtTGCT
48	SA6	184 GCGGTCACGGCTCCTCTTCGGAGGGCCGTTGAtTACTTgGCGGGaGGGGCcGCCCTgTGCT
45-50	consensus	GCGGTCACGGCTCCTCTTCGGAGGGcCGTTGACTACTtTaGCGGGaGGGgGCTGCCCTcTGCT
45	SA1	245 CCGCACTATACGTCCGcGACGCGTGCGGGGCAGTGTTtcTGGTAGGCCAAATGTTACCTA
47	SA5	245 CCGCACTATACGTCCGGGACGCGTGCGGGGCAGTGTTcTTGGTAGGCCAAATGTTACCTA
49	SA7	245 CCGCgCTATACGTCCGGGACGCGTGCGGGGCAGTGTTTTTGGTAGGCCAgATGTTcAgCTA
46	SA4	245 CCGCaCTATACGTCCGGGACGCGTGCGGGGCAGTGTTTTTGGTAGGCCAAATGTTACCTA
50	SA13	245 CCGCGTTATACGTCCGAGACGCGTGCGGGGCAGTGTTTTTGGTAGGtCAAATGTTACCTA
48	SA6	245 CCGCGTTATACGTCCGAGACGtGTGCGGGCAtTGTTTTTGGTAGGcCAAATGTTACCTA
45-50	consensus	CCGC-cTATACGTCCGgGACGcGTGCGGGCAGTGTTtcTGGTAGGcCAaATGTTACCTA
45	SA1	306 TAGGCCTCGCCAGCATAcAcAGTGcAGGACTGCAACTGTTCCATTtTACAGtGGCCATATC
47	SA5	306 TAGGCCTCGCCAGCATACTACGGTGCAGGACTGCAACTGTTCCATTtTACAGcGGCCATATC
49	SA7	306 TAGGCCTCGCCAGCACACTACGGTGCAGGACTGCAACTGTTCCATTtTACAGTGGCCATATC
46	SA4	306 TAGGCCTCGCCAGCACACTACGGTGCaAGACTGCAAtTGcTcTATTtTACAGTGGCCATATC
50	SA13	306 TAGcCCTCGCCgGCATAaTgttGTGcAGGACTGCAACTGtTCCATTtTACAGTGGCCaATC
48	SA6	306 TAGgCCTCGCCaGCATgctAcgGTaCAGGACTGCAACTGcTCCATTtTACAGTGGCCaATC
45-50	consensus	TAGgCCTCGCCaGCAtactacgGTgCAGgACTGCAAcTGtTcCATTtTACAGtGGCCaATC
45	SA1	367 ACCGGCCACCGgATGGCtTGGGACATGATGATGAATTGGTCACCTACGACAGCCTTGcTGA
47	SA5	367 ACCGGCCACCGAATGGCATGGGACATGATGATGAATTGGTCACCTACGACAGCCTTGGTGA
49	SA7	367 ACCGGCCACCGAATGGCATGGGACATGATGATGAATTGGTCACCTACGACAGCCTTGGTGA
46	SA4	367 ACCGGCCACCGATGGCATGGGACATGATGATGAATTGGTCACCTACGACgGCCTTGcTGA
50	SA13	367 ACCGGCCACCGATGGCATGGGACATGATGATGAATTGGTCACCTACaACAGCtTGGTGA
48	SA6	367 ActGGCCACCGATGGCATGGGACATGATGATGAATTGGTCACcCgCgACAGCtTGGTGA
45-50	consensus	ACcGGCCACCGgATGGCaTGGGACATGATGATGAATTGGTCACCTaCgACaGCcTTGGTGA



26/89

FIGURE 1G

<u>SEQ ID NO:</u>	<u>Isolate</u>	
45	SA1	428 TGGCCCAGaTGCTACGGATcCCCCAgGTGGTCATaGACATCATaGCCGGGGGCCACTGGGG
47	SA5	428 TGGCCCAGgTGCTACGGATTCCCCAaGTGGTCATtGACATCATTGCCGGGGGCCACTGGGG
49	SA7	428 TGGCCCAGTTGCTACGGATTCCCCAGGTGGTCATCGACATCATTGCCGGGGGCCACTGGGG
46	SA4	428 TGGCCCAGTTGCTACGGATTCCCCAGGTGGTCATCGACATCATTGCCGGGGGCCACTGGGG
50	SA13	428 TGGCCCAGTTGtTACGGATTCCCCAGGTGGTCATTGACATCATTGCCGGGGcCCACTGGGG
48	SA6	428 TGGCCCAaaTGcTACGGATTCCCCAGGTGGTCATTGACATCATTGCCGGGGgCCACTGGGG
45-50	consensus	TGGCCCAGtTGcTACGGATtCCCCAgGTGGTCATtGACATCATtGCCGGGGgCCACTGGGG
<u>SEQ ID NO:</u>	<u>Isolate</u>	
45	SA1	489 GGTCTTGTTtGCCGcCGCATACTTtGCGTCgGCcGCcAACTGGGCTAAGGTaGTGCTGGTt
47	SA5	489 GGTCTTGTTTCGCCGtCGCATACTTCGCGTCAGCGGCTAACTGGGCTAAGGTTGTGCTGGTC
49	SA7	489 GGTCTTGTTTCGCCGCGCATATTTCGCGTCAGCGGCTAACTGGGCTAAGGTTGTGCTGGTC
46	SA4	489 GGTCTTGTTtGCCGCGCATATTTCGCGTCAGCGGCTAACTGGGCTAAGGTTaTaCTGGTC
50	SA13	489 GGTCTTGTTTCGCCGCGCATACTaCGCGTCGGCGGCTAACTGGGcAAGGTTGTGCTGGTC
48	SA6	489 GGTCTTGTTTCGCCGCGtGCATACTtCGCGTCGGCGGCTAACTGGGcLAAGGTTGTGCTGGTC
45-50	consensus	GGTCTTGTTcGCCGccGCATACTtCGCGTC-GCgGCTAACTGGGCTAAGGTTgTgCTGGTc
<u>SEQ ID NO:</u>	<u>Isolate</u>	
45	SA1	550 CTGTTcCTGTTTGCGGGGGTCGATGGC
47	SA5	550 CTGTTTCTGTTTGCGGGGGTCGATGGC
49	SA7	550 TTGTTTCTGTTTGCGGGGGTCGATGCC
46	SA4	550 TTGTTTCTGTTTGCGGGGGTCGATGCC
50	SA13	550 cTGTTTCTGTTTGCGGGGGTCGATGCC
48	SA6	550 tTGTTTCTGTTTGCGGGGGTtGATGCC
45-50	consensus	-TGTTtCTGTTTGCGGGGGTcGATGcC

27/89

FIGURE 1B

SEQ ID NO:	Genotype	
30-33	(IV/2b)	1 GTGGAAGTcAGgAACAtCAGTTctAGcTACTAcGCCACCAATGATTGCTCaaACAaCAGCA
34	(2c)	1 GTGGAGGTCAAGGACACCGGCGACTCCTACATGCCGACCAACGATTGCTCCAATCTTAGTA
26-29	(III/2a)	1 GcccAAGTGAgaAACACCAgtacCaGcTACATGGTGACcAAcGACTGtTCCAAAtGacAGCA
35-39	(V/3a)	1 cTAGAGTGGCGGAATacGTctGGCCTCTAtgTCCTcACCAACGACTGTtCCAATAGCAGTA
9-25	(II/1b)	1 tAtGAaGTGCgCAACGTgTCCGGGgtgTaccAtGTcACgAAcGACTGcTCCAATCaaAGca
1-8	(I/1a)	1 tACCAAGTgCGCAACTCcaCgGGgCTtTACCATGTcACCAATGAAtTGCCTTAACTCGAGtA
40	(4a)	1 GAGCACTACCGGAATGCTTCGGGcATCTATCACATCACCaaTgATTGTCCGAATTCCAGTA
42-43	(4c)	1 GTtAACTATCgCAATGCCTCGGGCGTCTATCACgTCACCAACGACTGCCCGAACTCGAGCA
44	(4d)	1 TACAACCTATCGCAACAGCTCGGGTGTCTACCATGTcACCAACGATTGCCCGAACTCGAGCA
41	(4b)	1 GTGCACTACCGGAATGCTTCGGGCGTCTATCATGTcACCAATGATTGCCCTTAACACAGCA
45-50	(5a)	1 GTtCCcTACCGaaAtGCCTctGGGGTtTAtCATGTcACCAATGAAtTGCcCAAACTCtTCCA
51	(6a)	1 CTTACCTACGGCAACTCCAGTGGGCTATACCATCTCACAaATGATTGCCcCAACTCCAGCA
1-51	consensus	A TA AC AA GA TG C AA
SEQ ID NO:	Genotype	
30-33	(IV/2b)	62 TCACCTGGCAaCTCACCaaCGCAGTtCTCCACCTTCCCGGATGCGTCCCaaTGTGAGAATGA
34	(2c)	62 TCGTTTGGCAGCTTGAAAGGAGCAGTGCTTCATACCTCTGGATGCGTCCCTTGTGAGCGTAC
26-29	(III/2a)	62 TCACcTGGCAaCTcAgGCcGCGGTcCTCCACGTcCCCGGGTgTgTCCGTGcGAGaaagt
35-39	(V/3a)	62 TtGTGTATGAGGCCGATGACGTcATTCTGCACACACCTcGGCTGTGTACCTTGTGTTCAGGA
9-25	(II/1b)	62 TtGTGTatGAggCagcgGACaTGATcaTGCACaCccCGGgTGcgTgCCCTGcGTtCgGGA
1-8	(I/1a)	62 TtGTGTACGAGgCgGCcGATgCcATcCTgCacaCtCCgGGgTGTGTcCTTTCGCTTCGcGA
40	(4a)	62 TAGTCTATGAAGCTGACCATCAGATCTTACACTTGGCGGGGTGCGTACCTTGTGTGATGAC
42-43	(4c)	62 TAGTGTATGAGGCCGAACACCAgATCtTACACCTCCcAGGGTGTcTgCCCTGTGTGAGGGt
44	(4d)	62 TAGTCTATGAAACCGATTACCACATCTTACACCTCCCGGGATGCGTTCCTTGCCTGAGGGA
41	(4b)	62 TAGTGTACGAGACGGAGCACCACATCATGCACTTGCCAGGGTGTGTCCCTTGTGTGCGGAC
45-50	(5a)	62 TaGTcTAcGAGGCTGAtaaCCTGATctTgCacGCACCTGGtTGCCTGCCcTGTGTcaggcA
51	(6a)	62 TCGTGCTGGAGGCGGATGCTATGATCTTGcATTGTGCTGGATGCTTGCTTGTGTGAGGGT
1-51	consensus	T A T T CA CC GG TG T CC TG G
SEQ ID NO:	Genotype	
30-33	(IV/2b)	123 cAATGGCACCCcTGcGTGCTGGATACAAGTgACACCTAATGTGGCTGTGAAACACCGcGGC
34	(2c)	123 CGCCAACGTCTCTCGATGTTGGGTGCCGGTTCGCCCAATCTCGCCATAAGTCAACCTGGC
26-29	(III/2a)	123 gGGAAAtaCaTctCGgTGCTGGATACCGGTctCaCCAAAcGTgGCcGTGCaGcAGCCcGGC
35-39	(V/3a)	123 CGGcAATACATCcAcGTGCTGGACCCcCaGTGACaCCTACaGTGGCAGTCAGGTACGTCCGA
9-25	(II/1b)	123 gaacAActcCTCcgcgTgCTGGGTaGCGCTcaCtCCcACgCTcGCgGCcAGGAacgcccAgC
1-8	(I/1a)	123 GGgTaaCgcctCGAggTGTGGGTGgCGgTGaCCCCCAGcGTgGCCACcAGGGAcGGCAaa
40	(4a)	123 TGGGAACACATCGCGTTCCTGGAGCGCCGGTGACGCCTACAGTGGCTGTGCGCACACCCGGGC
42-43	(4c)	123 tGGGAAtCAGTCACGCTGCTGGGTGGCCCTTACTCCcACCGTGGCGGtGtCTTATATCGGT
44	(4d)	123 AGGGAACAAGTCTACATGCTGGGTGTCTCTACCCCCACCGTGGCTGCGCAACATCTGAAT
41	(4b)	123 GGAGAATACTTCTCGCTGCTGGGTGCCCTTGACCCCCACTGTGGCCCGGCCCTATCCCAAC
45-50	(5a)	123 agaTAATGTcAGTAggTGCTGGGTcCAaATCACCCCCACatTgTCAGCCCCGAaaccTCGGA
51	(6a)	123 CGATGATCGGTCCACCTGTTGGCATGCTGTGACCCCCACCTGGCCATACCAaATGCTTCC
1-51	consensus	TG TGG T C CC A T C

28/89

FIGURE 1E

SEQ ID NO: Genotype  
 30-33 (IV/2b)  
 34 (2c)  
 26-29 (III/2a)  
 35-39 (V/3a)  
 9-25 (II/1b)  
 1-8 (I/1a)  
 40 (4a)  
 42-43 (4c)  
 44 (4d)  
 41 (4b)  
 45-50 (5a)  
 51 (6a)

184 GCaCTcACTCACAACCTGCGAaCaCatgTcGacaTGATcGTAATGGCAGCTACGGTCTGCT  
 184 GCTCTCACTAAGGGCCTGCGAGCACACATCGATATCATCGTGTATGCTCTACGGTCTGTT  
 184 GCcCTcACGCAGGGCTTGCGGACgCACATcGACATGGTtGTGATGTCCGCCACGCTCTGCT  
 184 GCAACCACCGCtTCGATACGCAGTCATGTGGACCTatTaGTGGGCGCGGCCACgaTGTGCT  
 184 gTCCcCACTAcGaCaATACGACgCacGTGATtTGCTCGTTGGGGCGGCTgetTTCTGcT  
 184 CTCCCcgCAaCGCAGCTtCGACGTcACATCGATCTGCTtGTcGGgAGcGCCACCCTCTGcT  
 184 GCTCCGCTTGAGTCTGTTCCGGGACATGTGGACTTAATGGTAGGCGCGGCCACTTTGTGTT  
 184 GCtCCGCTTGACtCCcTCCGGAGACATGTGGACCTGATGGTgGGCGCcGCTACTGTaTGCT  
 184 GCTCCGCTTGAGTCTTTGAGACGTcACGTGGATCTGATGGTGGGCGCGGCCACTCTCTGCT  
 184 GCACCCTTAGAGTCTCGCAGGCATGTAGACCTGATGGTGGGTGCGGCTACTATGTGTT  
 184 GCGGTcACGGCTCTCTCTCGGAGGGcCGTTGAcTAcTtAGCGGGaGGgGCTGCCCTCTGCT  
 184 ACGCCCGCAACGGGATTCCGCGAGGCATGTGGATCTTCTTGCGGGCGCGCAGTGGTTTGCT

1-51 consensus

T G T GA T G GC T TG T

SEQ ID NO: Genotype  
 30-33 (IV/2b)  
 34 (2c)  
 26-29 (III/2a)  
 35-39 (V/3a)  
 9-25 (II/1b)  
 1-8 (I/1a)  
 40 (4a)  
 42-43 (4c)  
 44 (4d)  
 41 (4b)  
 45-50 (5a)  
 51 (6a)  
 1-51 consensus

245 CGGCCTTGATGTGGGAGACgTgTgCGGGGCGGTGATGATcGtGTGCGAGGCTtTCATAaT  
 245 CTGCCCTTTATGTGGGGGACGTGTGTGGCGCGCTGATGCTGGCCGCTCAGGTCTGTCGTCT  
 245 CcGCTcTtTACGTGGGGGAcCTCTCGGGcGGGgTgATGCTCGCaGCcAgATGTTcATTtgt  
 245 CTGCGCTCTACGTGGGtGATaTGTGTGGGGCCGTCTTtCTcGTGGGACAGCCTTcACGTT  
 245 CCGctATGTAcGTGGGgGAtCTcTGCGGaTCTGTTtTCTCTegTCTCcCAGcTGTTCACctT  
 245 CGGCCCTCTAcGTGGGGGACtTGTCGGGTCTGTCTTtCTtGTCTgGtCAaCTGTTcACctT  
 245 CTGCCCTCTATGTGTGGGACCTCTCGGAGGTGcCTTCTCTGATGGGCGAGATGATCACTTT  
 245 CtGCCCTCTACgTGTGGaGAtCTGTGCGGTGGtGcATTCTTGGTTGGcCAGATGTTtTCcTT  
 245 CCGCCCTCTACATCGGAGACGTGTGTGGGGGTGTGTCTTGGTCCGTCAACTGTTcACCTT  
 245 CCGCCTTCTACATTGGAGATCTGTGTGGAGCGTCTTCTAGTGGGCGCAGTGTTCGACTT  
 245 CCGCgCTATACGTGGGgGACGcGTGCGGGGCAgTGTtTtTGGTAGGcCAaATGTTcACcTA  
 245 CATCCCTGTACATCGGGGACCTGTGTGGCTCTCTCTTTTGGCGGGACAACtATTcACCTT  
 C T TA T GG GA TG GG T T CA T

SEQ ID NO: Genotype  
 30-33 (IV/2b)  
 34 (2c)  
 26-29 (III/2a)  
 35-39 (V/3a)  
 9-25 (II/1b)  
 1-8 (I/1a)  
 40 (4a)  
 42-43 (4c)  
 44 (4d)  
 41 (4b)  
 45-50 (5a)  
 51 (6a)  
 1-51 consensus

306 ATCGCCaGAACgCCACaACTTtACCCaAGAGTGCAACTGTTCCATCTACCAAGGTcAtATC  
 306 GTCGCCACAACACCATACGTTTGTCCAGGAATGCAACTGTTCCATATACCGGGCCGCAATT  
 306 CTCGCCCaCaCaCAGTgTtTGTGCaAGaATGCAAtTGCTCAtcTACCTtGGtACCATC  
 306 CAGACCTcCGTCGCCATCAAAcGTCCAGACCTGTAACGTCTCGCTGTACCCAGGCCAcCTT  
 306 cTCgCTcGcCggcAtgaGACagtaCAGgActGcAAcTGcTCAaTCTATCCcGGcCagTa  
 306 cTCTcCCAGgCgCCaCTGGACaACGCaAGaCTGcAAcTGTTCtATCTAtCCcGGCCAtATa  
 306 TCGGCCGCGTCCGCACTGGACCACGCAGGAGTGCAATGTTCATCTACACTGGCCATATC  
 306 CCAGCCGCGACGCCACTGGACTACGCAGGACTGCAATGTGTTCTATCTAcGCaGGGCaTaTc  
 306 CCAACCTCGCCGCCACTGGACCACCCaAGACTGCAATGTTCATCTACACAGGACATATC  
 306 CCGACCcGCGCGCACTGGACCACCCAGGATTGCAACTGCTCCATCTATCTGTTcACGTC  
 306 TAGgCTTCGCCCaGCAtactacgGTgCagGACTGCAAcTGtTCcATTtACAGtGGCCAtATC  
 306 TCAGCCCGCGCTcATTGGACTGTGCAAGACTGCAACTGCTCCATCTATACAGGCCACGTC  
 CC C CA TG AA TG TC T TA GG T

SEQ ID NO: Genotype  
 30-33 (IV/2b)  
 34 (2c)  
 26-29 (III/2a)  
 35-39 (V/3a)  
 9-25 (II/1b)  
 1-8 (I/1a)  
 40 (4a)  
 42-43 (4c)  
 44 (4d)  
 41 (4b)  
 45-50 (5a)  
 51 (6a)  
 1-51 consensus

367 ACCGGCCACCGCATGGCaTGGGACATGATGCTaAACTGGTCACCAACTCTtACCATGATCC  
 367 ACGGGACACCGCATGGCTTGGGATATGATGATGAACCTGGTCGCCCACTACCACCATGCTCC  
 367 ACTGGaCACCgTATGGCATGGGACATGATGATGAACCTGGTCGCCCAcGgCCACcaTGATCc  
 367 TCAGGACATCGaATGGCTTGGGATATGATGATGAATGGTCCCCCGctGTGGGTATGGTGG  
 367 tCAGGTcACCGcATGGcTGGGaLATGATGATGAACCTGGTCACCTcACAgCaGCcctAGTg  
 367 ACGGGtCAcCGcATGGCaTGGGATATGATGATGAACCTGGTCGCCCAcGgCCAGcCTGGTg  
 367 ACCGGCCACAGGATGGCGTGGGACATGATGATGAACCTGGAGCCCTACCACCACTCTGCTCC  
 367 ACgGGCCACAGgATGGCATGGGACATGATGATGAACCTGGAGTCCCAcAACCCACCTGctTc  
 367 TCGGGCCACAGGATGGCCTGGGACATGATGATGAACCTGGAGCCCTACCAGCCGCTGATTa  
 367 ACcGGCCACCGgATGGCaTGGGACATGATGATGAATGGTCACCTcAcgACaGCcctTgTGA  
 367 ACCGGCCACAGGATGGCTTGGGACATGATGATGAACCTGGTCACCCACAACCACTCTGGTCC  
 C GG CA G ATGGC TGGGA ATGATG T AA TGG CC C T T

29/89

FIGURE 1H

SEQ ID NO:	Genotype	
30-33	(IV/2b)	428 TcGCCTAtGCcGCtCGTGTtCCTGAgCTAGtCCTtgAaGTTtGTCTTCGGcGGcCATTGGGG
34	(2c)	428 TGGCGTACTTGGTGCGCATCCCCGAAGTCATCTTGGATATTGTTACAGGAGGTCAATTGGGG
26-29	(III/2a)	428 TGGCGTACGcGATGCGCGTtCCCCGAGGTCAATaTAGACATCaTtaGCCGgGCTcCACTGGGG
35-39	(V/3a)	428 TgGCGCACgTcCTGCGtTtTGCCCCAGACCtTGTTcGACATAaTaGCCcGGGGCCCAATTGGGG
9-25	(II/1b)	428 TaTCGAgTtTaCTCCGgaTCCCaCAAGCTgTcGtGGAcATGGTggCgGGgGCCCACTGGGG
1-8	(I/1a)	428 TaGCtCAGCTGCTCcgGaTCCCGCaAGCCaTCTTGGAcATGATCGCTGGtGCcCACTGGGG
40	(4a)	428 TCGCCcAGATCATGAGGGTCCCCACAGCCTTTCTCGACATGGTTGCCGAGGCCACTGGGG
42-43	(4c)	428 TCGCCcAGGTcATGAGGATCCCTAGCACTCTGGTaGatCTACTCgCTGGAGGGCACTGGGG
44	(4d)	428 TCGCCcAACTTATGAGGATCCcAGGCGCCATGGTTCGACCTGCTTGCAGGCGGGCACTGGGG
41	(4b)	428 TGGCTCAGATCTTACGGATCCCCAGTCTATCTAGGTGACTTGTcACCGGGGGTCACTGGGG
45-50	(5a)	428 TGGCCcAgTtGCTACGGATtCCCCAgGTGGTCATtGACATCATtGCCGGGGgCCACTGGGG
51	(6a)	428 TATCTAGCATCTTgAGGGTACCTGAGATTGTGCGAGTGTGATATTTGGTGGCCATTGGGG
1-51	consensus	T C G T CC T T GG G CA TGGGG
30-33	(IV/2b)	489 cGTGGTGTtTGGCTTGGCCTATTtCTCCATGCAGGGAGCGTGGGCCAAaGTCATtGCCATC
34	(2c)	489 TGTAAATGTtTGGCCTCGCTTACTTCTCCATGCAGGGATCGTGGGCGAAGGTCACTGTTATC
26-29	(III/2a)	489 CGTCaTGTTcGGCtTaGCCTACTTCTCTATGCAGGGAGCGTGGGCGAAaGTCgTTGTCACT
35-39	(V/3a)	489 CATCtTGGCgGGCCTAGCCTATTaTcATGCAGGGCAACTGGGCCAAGGTCTGCTATcATC
9-25	(II/1b)	489 agTCCTgGCGGGCCTtGCcTACTaTtCCATGGtGGgAACTGGGCTaAGGTtTgATTGTg
1-8	(I/1a)	489 AGTCCTaGCGGGCATAGCGTATTtTCCATGGtGGgAACTGGGCGAAGGTCTgTgTaGTg
40	(4a)	489 CGTCCTCGCGGGCTTGGCGTACTTcAGCATGCAAGGCAATTGGGCCAAGGTAGTCTGGTC
42-43	(4c)	489 cgTCCtTgTtGGGtTGGGcGTACTTcAGtATGCAAGCTAATTGGGCCAAaGTCATcCTGGTC
44	(4d)	489 CATTCGGTTGGCATAGCGTACTTcAGCATGCAAGCTAATTGGGCCAAGGTATCTGGTC
41	(4b)	489 AGTTCTTGCTGGTCTAGCTTTCTTcAGCATGCAGAGTAaCTGGGCGAAGGTCACTGGTC
45-50	(5a)	489 GGTCTTGTtGCGGcCGCATAcTtCGCGTCgGCgGCTAACTGGGCTaAGGTtGtCTGGTC
51	(6a)	489 GATACTACTAGCCGTtGGCTACTTTGGCATGGCTGGCAACTGGCTAAAGATTCTGGCTGTT
1-51	consensus	T T G GC T T TGG AA GT T
30-33	(IV/2b)	550 CTCCTtCTTGTcGCAGGAGTGGAtGCA
34	(2c)	550 CTCCTGCTGACTGCTGGGGTGGAGGCG
26-29	(III/2a)	550 CTtTtTGCTggCcGCTGGgGTGGACGCG
35-39	(V/3a)	550 ATGgTTATGTTTTcAGGgGTcGAtGCC
9-25	(II/1b)	550 aTGCTACTcTTTGCcGGcGTTcGAcGGg
1-8	(I/1a)	550 CTGtTGCTgTtTgCCGGCGTcGAtGCG
40	(4a)	550 CTTTTCTCTTTTGGCTGGGGTAGACGCC
42-43	(4c)	550 CTTTTCTCTtCGCTGGAGTTGATGCC
44	(4d)	550 CTGTTTCTCTTTTGGCTGGAGTcGACGCT
41	(4b)	550 CTATTCTCTTTTGGCGGGGTcGAGGGA
45-50	(5a)	550 tTGTTtCTGTTTTGGGGGGTcGATGcC
51	(6a)	550 CTGTTCTATTtGCAGGGGTGAAGCA
1-51	consensus	T T T C GG GT GA G

30/89

FIGURE 2A

SEQ ID NO:	Isolate	
56	S14	1 YQVRNSTGLYHVTNDCPNSSIVYEADAILHAPGCVPCVREGNtSRCWVAMTPTVATRDGK 
52	DK7	1 YQVRNSTGLYHVTNDCPNSSIVYEADAILHTPGCVPCVREGNvSRCWVAMTPTVATRDGK 
59	US11	1 YQVRNSTGLYHVTNDCPNSSIVYEADAILHTPGCVPCVREGNaSRCWVAMTPTVATRDGK 
55	DR4	1 HQVRNSTGLYHVTNDCPNSSIVYEADAILHTPGCVPCVREGNtSRCWVAVTPTVATRDGK 
54	DR1	1 HQVRNSTGLYHVTNDCPNSSIVYEADAILHAPGCVPCVREGNAsRCWVAVTPTVATRDGK 
53	DK9	1 YQVRNSSGLYHVTNDCPNSSIVYEADAILHSPGCVPCVREGNASKCWVAVAPTPTVATRDGK 
58	SW1	1 YQVRNSSGLYHVTNDCPNSSIVYETADAILHSPGCVPCVREGdApKCWVAVAPTPTVATRDGK 
57	S18	1 YQVRNSTGLYHVTNDCPNSSIVYETADtILHSPGCVPCVREGnAsrCWVpVAPTPTVATRDGK 
52-59	consensus	yQVRNSTGLYHVTNDCPNSSIVYEaADaILH-PGCVPCVREGnaSrCWVavtPTVATRDGK
56	S14	62 LPatQLRRyIDLLVGSATLCSALYVGDLGGSVFLVGQLFTFSPPRRlWTTQdCNCsIYPGHI 
52	DK7	62 LPTaQLRRHIDLLVGSATLCSALYVGDLGGSVFLVGQLFTFSPPRRHWTQGCNCsIYPGHI 
59	US11	62 LPTTQLRRHIDLLVGSATLCSALYVGDLGGSVFLVGQLFTFSPPRRHWTQGCNCsIYPGHI 
55	DR4	62 LPTTQLRRHIDLLVGSATLCSALYVGDLGGSVFLVGQLFTFSPPRhWTTQdCNCsIYPGHI 
54	DR1	62 LPTTQLRRHIDLLVGSATLCSALYVGDLGGSVFLVGQLFTFSPPRRHWTQdCNCsIYPGHI 
53	DK9	62 LPATQLRRHIDLLVGSATLCSALYVGDLGGSVFLVGQLFTFSPPRRHWTQdCNCsIYPGHI 
58	SW1	62 LPATQLRRHIDLLVGSATLCSALYVGDLGGSVFLVSQLFTFSPPRRHWTQdCNCsIYPGHI 
57	S18	62 LPATQLRRHIDLLVGSATLCSALYVGDLGGSVFLVSQLFTiSPRRHWTQdCNCsIYPGHI 
52-59	consensus	LP-tQLRRhIDLLVGSATLCSALYVGDLGGSVFLVgQLFTfSPRRhWTTQdCNCsIYPGHI
56	S14	123 TGHrMAWdMMNWSPTTALVVAQLLRIPOAILDMLAGAHWGVLAGIAYFSMVGNWAKVLVV 
52	DK7	123 TGHrMAWdMMNWSPTTALVVAQLLRIPOAILDMLAGAHWGVLAGIAYFSMVGNWAKVLVV 
59	US11	123 TGHrMAWdMMNWSPTaALVVAQLLRIPOAILDMLAGAHWGVLAGIAYFSMVGNWAKVLVV 
55	DR4	123 TGHrMAWdMMNWSPTTALVVAQLLRIPOAILDMLAGAHWGVLAGIAYFSMVGNWAKVLVV 
54	DR1	123 TGHrMAWdMMNWSPTTALVMAQLLRIPOAILDMLAGAHWGVLAGIAYFSMVGNWAKVVVV 
53	DK9	123 TGHrMAWdMMNWSPTaALVMAQLLRIPOAILDMLAGAHWGVLAGIAYFSMVGNWAKVVVV 
58	SW1	123 TGHrMAWdMMNWSPTTALVvAQLLRIPOAVLDMLAGAHWGVLAGIAYFSMVGNWAKVLiV 
57	S18	123 TGHrMAWdMMNWSPTTALViAQLLRvPOAVLDMLAGAHWGVLAGIAYFSMaGNWAKVLiV 
52-59	consensus	TGHrMAWdMMNWSPTtALVvAQLLRiPOaILDMLAGAHWGVLAGIAYFSMvGNWAKVlVV

31/89

FIGURE 2A

<u>SEQ ID NO:</u>	<u>Isolate</u>	
56	S14	184 LLLFAGVDA 
52	DK7	184 LLLFAGVDA 
59	US11	184 LLLFAGVDA 
55	DR4	184 LLLFAGVDA 
54	DR1	184 LLLFAGVDA 
53	DK9	184 LLLFAGVDA 
58	SW1	184 LLLFAGVDA 
57	S18	184 LLLFAGVDA 
52-59	consensus	LLLFAGVDA

32/89

FIGURE 2B

SEQ ID NO:	Isolate	
75	T10	1 YEVRNVSGmYHVTNDCSNSSIVFEAdlIMHTPGCVPCVREgNsSRCWVALTPTLAARNtS 
62	DK1	1 YEVRNVSGvYHVTNDCSNSSIVYEAvdvIMHTPGCVPCVRENNhSRCWVALTPTLAARNAS 
64	HK4	1 hEVhNVSGiYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNAS 
76	US6	1 YEVRNVSGmYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNAS 
68	IND8	1 YEVRNVSGvYHVTNDCSNSSIVYEAADMIMHTPGCVPCVREGNfSsCWVALTPTLAARNAS 
67	IND5	1 YEVRNVSGvYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNAS 
73	SW2	1 YEVRNVSGvYHVTNDCSNSSIVYETADMIMHTPGCVPCVREaNSSRCWVALTPTLAARNtS 
63	HK3	1 YEVRNVSGiYHVTNDCSNSSvVYETADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNVS 
66	HK8	1 YEVRNVSGiYHVTNDCSNSSIVYETADMIMHTPGcmPCVRENNSSRCWVALTPTLAARNVS 
61	D3	1 YEVRNVSGVYqVTNDCSNSSIVYETADMIMHTPGCVPCVREdNSSRCWVALTPTLAARNsS 
74	T3	1 YEVRNVSGVYyVTNDCSNSSIVYETADMIMHTPGCVPCVREaNSSRCWVALTPTLAARNAS 
65	HK5	1 YEVRNVSGVYHVTNDCSNlSIVYETldMIMHTPGCVPCVRENNSSRCWVALaPTLAARNAS 
71	S45	1 YEVRNVSGaYHVTNDCSNSSIVYEAvdvIlHTPGCVPCVRENNSSRCWVALTPTLAARNSS 
72	SA10	1 YEVRNVSGmYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNSS 
69	P10	1 YEVRNVSGvYHVTNDCSNSSIVYEAADMIMHTPGCVPCVRENNSSRCWVALTPTLAARNSS 
60	D1	1 YEVRNVSGvYHVTNDCSNSSIVYEtADMIMHTPGCVPCVREdNSSRCWVALTPTLAARNgn 
70	S9	1 YEVRNVSGaYHVTNDCSNSSIVYEAAdvIMHTPGCVPCVqEgNSSqCWVALTPTLAARNat 
60-76	consensus	yEVrNVSGvYhVTNDCSNsSiVyEaaDmImHTPGCVPCVrEnNsSrCWVALtPTLAARNas 

33/89

FIGURE 2B

SEQ ID NO:	Isolate	
75	T10	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETIQDCNCSIYPGH1
62	DK1	62 IPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETaQDCNCSIYPGHV
64	HK4	62 IPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
76	US6	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
68	IND8	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
67	IND5	62 VstTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
73	SW2	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
63	HK3	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
66	HK8	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
61	D3	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQeCNCSIYPGHV
74	T3	62 VPTKTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
65	HK5	62 VPTTaIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
71	S45	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRHETVQDCNCSIYPGHV
72	SA10	62 VPTTTIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRyETVQDCNCSIYPGrV
69	P10	62 VPTTAIRRHVDLLVGAAAFCSAMYVGDLCGSVFLVSQFTFSPPRRhwTVQDCNCSIYPGHV
60	D1	62 VPTTAIRRHVDLLVGAAAFCSAMYVGDLCGSVFLISQFTFSPPRRHETVQeCNCSIYPGHV
70	S9	62 VPTTtIRRHVDLLVGAAAFCSAMYVGDLCGSVFLISQFTFSPPRRHETVQnCNCSIYPGHV
60-76	consensus	vpTttIRRHVDLLVGAAAFCSAMYVGDLCGSVfLvSQLFTFSPPRhetVQdCNCSIYPGHv



34/89

FIGURE 2B

SEQ ID NO:	Isolate		
75	T10	123	SGHRMAWDMMNWSPTTALVVSQLLRIPQAVvDMVtGAHWGVLAGLAYYSMAGNWAKVLIV
62	DK1	123	SGHRMAWDMMNWSPTTALV1SQLLRIPQAVvDMVAGAHWGVLAGLAYYSMAGNWAKVLIV
64	HK4	123	SGHRMAWDMMNWSPTAALVVSQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
76	US6	123	SGHRMAWDMMNWSPTAALVVSQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
68	IND8	123	SGHRMAWDMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGILAGLAYYSMVGNWAKVLIV
67	IND5	123	SGHRMAWDMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGILAGLAYYSMVGNWAKVLIV
73	SW2	123	SGHRMAWDMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
63	HK3	123	SGHRMAWDMMNWSPTAALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
66	HK8	123	SGHRMAWDMMNWSPTtALVVSQLLRIPQAIvDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
61	D3	123	TGHRMAWDMMNWSPTaALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
74	T3	123	TGHRMAWDMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
65	HK5	123	TGHRMAWDMMNWSPTTALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
71	S45	123	TGHRMAWDMMNWSPTaALVVSQLLRIPQAVVDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
72	SA10	123	TGHRMAWDMMNWSPTtALVVSQLLRIPQAIvDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
69	P10	123	sGHRMAWDMMNWSPTaALVVSQLLRIPQAI1DvVAGAHWGVLAGLAYYSMVGNWAKVLIV
60	D1	123	TGHRMAWDMMNWSPTTALVVSQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
70	S9	123	TGHRMAWDMMNWSPTTALVVSQLLRIPQAVMDMVAGAHWGVLAGLAYYSMVGNWAKVLIV
60-76	consensus		sGHRMAWDMMNWSPTaALVVSQLLRIPQAVvDmVaGAHWGvLAGLAYYSMVGNWAKVLIV

35/89

FIGURE 2B

<u>SEQ ID NO:</u>	<u>Isolate</u>	
75	T10	184 mLLFAGVDG 
62	DK1	184 lLLFAGVDG 
64	HK4	184 mLLFAGVDG 
76	US6	184 lLLFAGVDG 
68	IND8	184 MLLFAGVDG 
67	IND5	184 MLLFAGVDG 
73	SW2	184 MLLFAGVDG 
63	HK3	184 MLLFAGVDG 
66	HK8	184 MLLFAGVDG 
61	D3	184 MLLFAGVDG 
74	T3	184 lLLFAGVDG 
65	HK5	184 MLLFAGVDG 
71	S45	184 MLLFAGVDG 
72	SA10	184 MLLFAGVDG 
69	P10	184 MLLFAGVDG 
60	D1	184 MLLFAGVDG 
70	S9	184 MLLFAGVDG 
60-76	consensus	mLLFAGVDG

36/89

FIGURE 2C

SEQ ID NO:	Isolate	
77	T2	1 AQVrNTsrgYMTNDCSNeSITWQLQAAVLHVPGCiPCEr1GNTSRCWIPVtPNVAVRQPG 
78	T4	1 AQVrNTttnSYMTNDCSNDsITWQLQAAVLHVPGCVPCEktGNTSRCWIPVSPNVAVRQPG 
79	T9	1 AeVrNTSTSYMTNDCSNDsITWQLQAAVLHVPGCVPCErVGNsSRCWIPVSPNVAVRQPG 
80	US10	1 vqVrNTSTSYMTNDCSNDsITWQLaAAVLHVPGCVPCEkVGNtSRCWIPVSPNVAVRQPG 
77-80	consensus	aqVrNTstSYMTNDCSNDsITWQLaAAVLHVPGCvPCE-vGNtSRCWIPVsPNVAV--PG
SEQ ID NO:	Isolate	
77	T2	62 ALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMFIVSPrrHWfVQeCNCsIYPGTI 
78	T4	62 ALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMFIVSPQHfVQdCNCsIYPGTI 
79	T9	62 ALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMF1iSPQHfVQeCNCsIYPGTI 
80	US10	62 ALTQGLRTHIDMVMSATLCSALYVGDfCGGmMLAAQMF1vSPrHhFVQeCNCsIYPGTI 
77-80	consensus	ALTQGLRTHIDMVMSATLCSALYVGDLCGGVMLAAQMF1vSP-hHwFVQeCNCsIYPGTI
SEQ ID NO:	Isolate	
77	T2	123 TGHrMAWDMMnWSPTATMILAYAMRVPEVIiDIigGAHWGVmFGLAYFSMQAWAKVIVI 
78	T4	123 TGHrMAWDMMnWSPTATMILAYAMRVPEVIiDIvSGAHWGVmFGLAYFSMQAWAKVVVI 
79	T9	123 TGHrMAWDMMnWSPTtTILAYAMRVPEVIiDIISGAHWGVmFGLAYFSMQAWAKVVVI 
80	US10	123 TGHrMAWDMMnWSPTAtIILAYvMRVPEVIiDIISGAHWGVlFGLAYFSMQAWAKVVVI 
77-80	consensus	TGHrMAWDMMnWSPTAtMILAYaMRVPEVIiDIISGAHWGVmFGLAYFSMQAWAKVVVI
SEQ ID NO:	Isolate	
77	T2	184 LLLAAGVDA 
78	T4	184 LLLAAGVDA 
79	T9	184 LLLtAGVDA 
80	US10	184 LLLaAGVDA 
77-80	consensus	LLLaAGVDA

37/89

FIGURE 2D

<u>SEQ ID NO:</u>	<u>Isolate</u>	
82	DK11	1 VEVNtSSSYATNDCSNsSITWQLTNAVHLPGCVPCENDNGTLHCWIOVTPNVAVKHRG 
83	SW3	1 VEVNtSSSYATNDCSNsSITWQLTNAVHLPGCVPCENDNGTLHCWIOVTPNVAVKHRG 
84	T8	1 VEVNtSfSYATNDCSNNSITWQLTNAVHLPGCVPCENDNGTLRCWIOVTPNVAVKHRG 
81	DK8	1 VEVNtSsSYATNDCSNNSITWQLTNAVHLPGCVPCENDNGTLRCWIOVTPNVAVKHRG 
81-84	consensus	VEVRN-SsSYATNDCSNsSITWQLTNAVHLPGCVPCENDNGTL-CWIOVTPNVAVKHRG
<u>SEQ ID NO:</u>	<u>Isolate</u>	
82	DK11	62 ALTHNLRAHIdMIVMAATVCSALYVGDvCGAVMIVSQAFIvSPEHhFTQECNCsIYQGhI 
83	SW3	62 ALTHNLRAHVdMIVMAATVCSALYVGDmCGAVMIVSQAFIISPERHNFTQECNCsIYQGRI 
84	T8	62 ALTHNLRTHVdVIVMAATVCSALYVGDVCGAVMIaSQAFIISPERHNFTQECNCsIYQGHI 
81	DK8	62 ALTHNLRTHVdVIVMAATVCSALYVGDVCGAVMIvSQALIISPERHNFTQECNCsIYQGHI 
81-84	consensus	ALTHNLR-HvD-IVMAATVCSALYVGDvCGAVMIvSQAFIISPERHNFTQECNCsIYQGHI
<u>SEQ ID NO:</u>	<u>Isolate</u>	
82	DK11	123 TGHrMAWdMMLNWSPTLTMI LAYAArVPELVLeVVFgGHwGVVFGlAYfSMQgAWAKVIAI 
83	SW3	123 TGHrMAWdMMLNWSPTLTMI LAYAArVPELVLeVVFgGHwGVVFGlAYfSMQgAWAKVIAI 
84	T8	123 TGHrMAWdMMLNWSPTLTMI LAYAArVPELVLeVVFgGHwGVVFGlAYfSMQgAWAKVIAI 
81	DK8	123 TGHrMAWdMMLNWSPTLTMI LAYAArVPELlLqVVFgGHwGVVFGlAYfSMQgAWAKVIAI 
81-84	consensus	TGHrMAWdMMLNWSPTLTMI LAYAArVPELVLeVVFgGHwGVVFGlAYfSMQgAWAKVIAI
<u>SEQ ID NO:</u>	<u>Isolate</u>	
82	DK11	184 LLLVAGVDA 
83	SW3	184 LLLVAGVDA 
84	T8	184 LLLVAGVDA 
81	DK8	184 LLLVAGVDA 
81-84	consensus	LLL VAGVDA

38/89

FIGURE 2E

<u>SEQ ID NO:</u>	<u>Isolate</u>	
86	DK12	1 LEWRNVSGLYVLTNDCsNSSIVYEADDVILHTPGCVPCVQDGNSTCWTSPVTPTVAVRYVG 
87	HK10	1 LEWRNVSGLYVLTNDCpNSSIVYEADDVILHTPGCVPCVQDGNSTCWTSPVTPTVAVRYVG 
88	S2	1 LEWRNTSGLYVLTNDCSNSSIVYEADDVILHTPGCVPCVQDGNSTCWTSPVTPTVAVRYVG 
90	S54	1 LEWRNTSGLYVLTNDCSNSSIVYEADDVILHTPGCVPCVQDGNSTCWTSPVTPTVAVRYVG 
89	S52	1 LEWRNTSGLYVLTNDCSNSSIVYEADDVILHTPGCVPCVQDGNSTmCWTSPVTPTVAVRYVG 
86-90	consensus	LEWRNTSGLYVLTNDCsNSSIVYEADDVILHTPGCVPCVQDGNSTcCWTSPVTPTVAVRYVG
<u>SEQ ID NO:</u>	<u>Isolate</u>	
86	DK12	62 ATTASIRSHVDLLVGAATMCSALYVGDvCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPGHL 
87	HK10	62 ATTASIRSHVDLLVGAATMCSALYVGDmCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPGHL 
88	S2	62 ATTASIRSHVDLLVGAATMCSALYVGDmCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPGHL 
90	S54	62 ATTASIRSHVDLLVGAATLCSALYVGDmCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPGHL 
89	S52	62 ATTASIRSHVDLLVGAATLCSALYVGDmCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPGHv 
86-90	consensus	ATTASIRSHVDLLVGAATmCSALYVGDmCGAVFLVGQAFTFRPRRHQTVQTCNCSLYPGHL
<u>SEQ ID NO:</u>	<u>Isolate</u>	
86	DK12	123 SGHRMAWDMMMNWSPAVGMVVAHVLRPQTLFDIAGAHWGImAGLAYYSMOGNWAKVAII 
87	HK10	123 SGHRMAWDMMMNWSPAVGMVVAHVLRPQTLFDIAGAHWGILAGLAYYSMOGNWAKVAII 
88	S2	123 SGHRMAWDMMMNWSPAVGMVVAHVLRPQTVFDIAGAHWGILAGLAYYSMOGNWAKVAII 
90	S54	123 SGHRMAWDMMMNWSPAVGMVVAHILRLPQTLFDIAGAHWGILAGLAYYSMOGNWAKVAII 
89	S52	123 SGHRMAWDMMMNWSPAVGMVVAHILRLPQTLFDIAGAHWGILAGLAYYSMOGNWAKVAIv 
86-90	consensus	SGHRMAWDMMMNWSPAVGMVVAHVLRPQTLFDIAGAHWGILAGLAYYSMOGNWAKVAII
<u>SEQ ID NO:</u>	<u>Isolate</u>	
86	DK12	184 MVMFSGVDA 
87	HK10	184 MVMFSGVDA 
88	S2	184 MVMFSGVDA 
90	S54	184 MIMFSGVDA 
89	S52	184 MIMFSGVDA 
86-90	consensus	MVMFSGVDA

39/B9

**FIGURE 2F**

94 27

93 26

**93-94 consensus (26)**

VNYrNASGVYHvTNDcPNSSiVYEAEHQILHLPGClPCVRvGNQSRcwALTPTVAveYIG

**94** **27**

93 26

93-94 сопвельс (26)

[illegible]

94                      27

93 26

**93-94 consensus (Z6)**

[illegible]

**94**                      **27**

93 26

**93-94 consensus (26)**

```

184 LFLyAGVDA
    ||| ||||
184 LFLFAGVDA
    LFLFAGVDA

```

40/89

**FIGURE 2G**

SEQ ID NO:	Isolate	
98	SA5	1 VPYRNASGVYHVTNDCPNSSIVYEADNLILHAPGCVPCVkegNVSRCWVQITPTLSAPNLG
100	SA7	1 VPYRNASGVYHVTNDCPNSSIVYEADNLILHAPGCVPCVRq <sup>n</sup> NVSRCWVQITPTLSAPNLG
97	SA4	1 VPYRNASGVYHVTNDCPNSSIVYEADNLILHAPGCVPCVRQDNVSkCWVQITPTLSAPNLG
96	SA1	1 VPYRNASGVYHVTNDCPNSSIVYEAD <sup>s</sup> LILHAPGCVPCVRQDNVSRCWVQITPTLSAPtfg
99	SA6	1 VPYRNASGVYHVTNDCPNSSIVYEADDLILHAPGCVPCVRkDNVSRCWVhITPTLSAPSLG
101	SA13	1 VPYRNASGVYHVTNDCPNSSIVYEADDLILHAPGCVPCVRqgNVSRCWVqITPTLSAPSLG
96-101	consensus	VPYRNASGVYHVTNDCPNSSIVYEAD <sup>s</sup> LILHAPGCVPCVRq <sup>d</sup> DNVSR <sup>c</sup> WVQITPTLSAPnlg

SEQ ID NO:	Isolate	
98	SA5	62 AVTAPLRRVVDYLAGGAALCSALYVGDA CGAVFLVGOMFtYRPRQHTTVQDCNCSIYSGHI 
100	SA7	62 AVTAPLRRRAVDYLAGGAALCSALYVGDA CGAVFLVGOMFtYRPRQHTTVQDCNCSIYSGHI 
97	SA4	62 AVTAPLRRRAVDYLAGGAALCSALYVGDA CGAVFLVGOMFtYRPRQHTTVQDCNCSIYSGHI 
96	SA1	62 AVTAPLRRRAVDYLAGGAALCSALYVGDA CGAVFLVGOMFtYRPRQHTTVQDCNCSIYSGHI 
99	SA6	62 AVTAPLRRRAVDYLAGGAALCSALYVGDA CGAVFLVGOMFtYRPRQHTTVQDCNCSIYSGHI 
101	SA13	62 AVTAPLRRRAVDYLAGGAALCSALYVGDA CGAVFLVGOMFtYsPRRhnvVQDCNCSIYSGHI 
96-101	consensus	AVTAPLRRRAVDYLAGGAALCSALYVGDA CGAVFLVGOMFtYrPRQHTTVQDCNCSIYSGHI

SEQ ID NO:	Isolate		
98	SA5	123	TGHRMAWDMMNNWSP <del>T</del> TALVMAQvLRIPQVV <del>I</del> IAGGHWGVLFavAYFASAANWAKVVLV
100	SA7	123	TGHRMAWDMMNNWSP <del>T</del> TALVMAQ <del>L</del> LRIPQVV <del>I</del> IAGGHWGVLF <del>A</del> AAAYFASAANWAKVVLV
97	SA4	123	TGHRMAWDMMNNWSP <del>T</del> TALLMAQ <del>L</del> LRIPQVV <del>I</del> IAGGHWGVLF <del>A</del> AAAYFASAANWAKV <del>i</del> LV
96	SA1	123	TGHRMAWDMMNNWSP <del>T</del> TALLMAQ <del>M</del> LRIPQVV <del>I</del> IAGGHWGVLF <del>A</del> AAAYFASAANWAKVVLV
99	SA6	123	TGHRMAWDMMNNWSP <del>a</del> TALVMAQ <del>M</del> LRIPQVV <del>I</del> IAGGHWGVLF <del>A</del> AAAYFASAANWAKVVLV
101	SA13	123	TGHRMAWDMMNNWSP <del>c</del> TALVMAQ <del>L</del> LRIPQVV <del>I</del> IAG <del>a</del> HWGVLF <del>A</del> AAAYyASAANWAKVVLV
96-101	consensus		TGHRMAWDMMNNWSP <del>c</del> TALvMAQ <del>L</del> LRIPQVV <del>I</del> IAGgHWGVLF <del>A</del> aAYFASAANWAKvVLV

41/89

FIGURE 2G

<u>SEQ ID NO:</u>	<u>Isolate</u>	
98	SA5	184 LFLFAGVDg 
100	SA7	184 LFLFAGVDA 
97	SA4	184 LFLFAGVDA 
96	SA1	184 LFLFAGVDg 
99	SA6	184 LFLFAGVDA 
101	SA13	184 LFLFAGVDA 
96-101	consensus	LFLFAGVDA



42/89

FIGURE 2H

SEO ID NO:	Genotype	
81-84	(IV/2b)	1 VEVrNiSsSYIATNDCSNnSITWQLTnAVLHLPgCvPCENDNGTLrCWIQVTPNVAVKHrg
85	(2c)	1 VEVKDTGDSYMPINDCSNssIVWQLEGAVLHTPGCVPCERTANVSRCWVPVAPNLAI SQPG
77-80	(III/2a)	1 aqVKNtstSYMTNDCSNdsITWQLqAAVLHVPgCvPCekvGntSRCWIPVvPNVAVqqPG
86-90	(V/3a)	1 LEWRNtSGLYvLTNDCSnssIVYEADDVILHTPGCVPCVDGNTStCWtPvTPTVAVRYVG
60-76	(II/1b)	1 yEVrNVSGvYhVTNDCSNsiVyEaaDmImHTPGCVPCVrEnNeSrCWVALtPTLAARNas
52-59	(I/1a)	1 yQVRNStGLYHVtNDcPNssIVYEaAdAILHsPGCVPCVREgnasrCWVavtPTVATRDGK
91	(4a)	1 EHYRNASGIYHITNDcPNssIVYEADHILHLPgCvPCVMtGNtSRCWTPTVPTVAVAHPG
93-94	(4c)	1 VNYrNASGVYHVtNDcPNssIVYEAEHqILHLPgC1PCVrVGNQSRCWVALTPTVAvsYIG
95	(4d)	1 YNYRNSSGVYHVtNDcPNssIVYETDYHILHLPgCvPCVREGNKStCWVSLTPTVAAQHln
92	(4b)	1 VHYRNASGVYHVtNDcPNtSIVYETEHIMHLPGCVPCVrTENTSRCWVPLTPTVAAPYPN
96-101	(5a)	1 VPYRNASGVYHVtNDcPNssIVYEADnLILHAPgCvPCVrqdNVSrCWVqITPTLSAPnlG
102	(6a)	1 LTYGNSSGLYHLTNDcPNssIVLEADAMILHLPgCLPCVrVDDrSTCWHAvtPTTLAI PNAS
52-102	consensus	Y TNDC N S H PGC PC CW P
SEO ID NO:	Genotype	
81-84	(IV/2b)	62 ALTHNLRLtHvDmIVMAATVCSALYVGdvCGAVMIvSQafIiSPERhNftQECNCsIYQghI
85	(2c)	62 ALTKGLRAHIDIIVMSATVCSALYVGdVCGALMLAAQVVVSPQHHTFVQECNCsIYPGRI
77-80	(III/2a)	62 ALTQGLRTHIDMVMSATLCSALYVGd1CGGvMLAAQMFivSPqhHwFVQeCNCsIYPGTI
86-90	(V/3a)	62 ATTASIRSHVDLLVGAATmCSALYVGdmcGAVFLVGQAFtFRPRRHQTvQTCNCsIYPGHI
60-76	(II/1b)	62 vpTtctIRrHVdLLVGAAaFCSaMYVGDLGGSVFLvSQLFTfSPRrhetvQdCNCsIYPghv
52-59	(I/1a)	62 LPatQLRRhIDLLVGSATLCSALYVGDLGGSVFLvGQLFTfSPRrhwTtQdCNCsIYPGHI
91	(4a)	62 APLESFRRHVDLMVGAATLCSALYVGDLGCGAFtMGQMfTFRPRRHWTtQECNCsIYTgHI
93-94	(4c)	62 APLdS1RRHVdLMVGAATVCSALYvGDLGCGaFLVGQMfSPQPRRHWTtQdCNCsIYAGHI
95	(4d)	62 APLES1RRHVdLMVGGATLCSALYIGdVCGGVFLVGQLFTfQPRRHWTtQdCNCsIYTgHI
92	(4b)	62 APLES1RRHVdLMVGAATMCSAFYIGdLGGVFLVGQLFDfRPRRHWTtQdCNCsIYPghv
96-101	(5a)	62 AVTAPLRRaVDYLAGGAALCSALYVGDaCGAvFLVGQMfYrPRghcttVQdCNCsIYSGHI
102	(6a)	62 TPATGFRRHVDLLAGAavVCSSLYIGDLGGSFLAGQLFTfQPRRHWTtQdCNCsIYTgHV
52-102	consensus	R D A CS Y GD CG Q P Q CNCs Y G
SEO ID NO:	Genotype	
81-84	(IV/2b)	123 TGHrMAWDMMLNWSPTLTmILAYAArVPELVLeVVFGGHwGVVFLGAYFSMQAWAKVIAI
85	(2c)	123 TGHrMAWDMMLNWSPTTmILAYLVRIPEVILDIVTGGHwGVVFLGAYFSMQGSWAKVIVI
77-80	(III/2a)	123 TGHrMAWDMMLNWSPTaTmILAYaMrVPEVIdDiSgAHwGVmFLGAYFSMQAWAKVvVI
86-90	(V/3a)	123 SGHrMAWDMMLNWSPAVGmVVAHVLRFPQTLFDIAGAHwGILAGLAYFSMQGNWAKVAII
60-76	(II/1b)	123 sGHrMAWDMMLNWSPTaALVvSQLLRiPQAvvDmVaGAHwGVLAGLAYFSMvGNWAKVLIV
52-59	(I/1a)	123 TGHrMAWDMMLNWSPTtALVvaQLLRiPQaILDMiAGAHwGVLAGIAYFSMvGNWAKVIV
91	(4a)	123 TGHrMAWDMMLNWSPTTTLtLLAQIMRVPTAFtLDMVAGGHwGVLAGLAYFSMQGNWAKVVLV
93-94	(4c)	123 TGHrMAWDMMLNWSPTTTLtLAQVMRIPSTLVdLLaGGHwGVLVGLAYFSMQANWAKVILV
95	(4d)	123 TGHrMAWDMMLNWSPTATLVLAQLMRIPGAMVDLLAGGHwGILVGIAYFSMQANWAKVILV
92	(4b)	123 SGHrMAWDMMLNWSPTSALIMaQILRIPsILGDLLTGGHwGVLAGLAFFSMQSNWAKVILV
96-101	(5a)	123 TGHrMAWDMMLNWSPTtALVmaQILRIPQVVIDIAGgHwGVLFaaAYfASAAWAKVvLV
102	(6a)	123 TGHrMAWDMMLNWSPTTTLVSSILRVPEICASVIFGGHwGILLAVAYFGMAGNWLKVLAV
52-102	consensus	GHRMAWDM MLNWS R P G HWG A W KV

43/89

FIGURE 2H

<u>SEO ID NO.</u>	<u>Genotype</u>		
81-84	(IV/2b)	184	LLL <sup>V</sup> AGVDA
85	(2c)	184	LLL <sup>T</sup> AGVEA
77-80	(III/2a)	184	LLL <sup>a</sup> AGVDA
86-90	(V/3a)	184	MvMPSGVDA
60-76	(II/1b)	184	mLLFAGVDG
52-59	(I/1a)	184	LLL <sup>Fa</sup> GVDA
91	(4a)	184	LFLFAGVDA
93-94	(4c)	184	LFLFAGVDA
95	(4d)	184	LFLFAGVDA
92	(4b)	184	LFLFAGVEG
96-101	(5a)	184	LFLFAGV <sup>a</sup> Da
102	(6a)	184	LFLFAGVEA
52-102	consensus	GV	

**FIGURE 3**

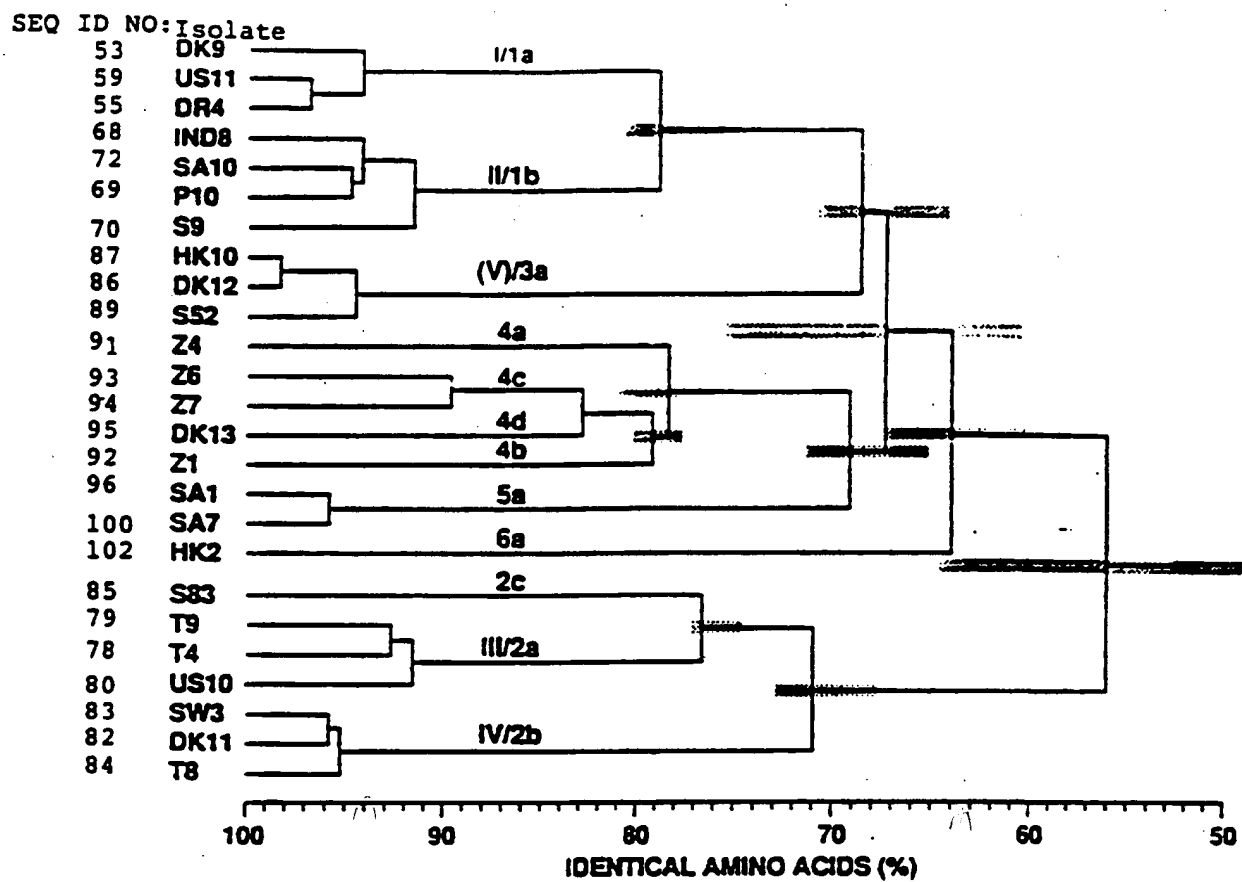
Genotype	SEQ ID NO: 52-102	Isolate	200	210	220	230	240	250	260	270	280
yeVRNvggVhVrTNDcNaSivyeada			lhtPGCvPCvregntarChvavtvtvaarnagaptttlRhnvdllvgatlcSalsvGd								
VEVRNtS8S-YA	82	DK11	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	83	SW3	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	84	DK8	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	85	T8	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	86	DK12	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	87	HK10	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	88	S2	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	89	S54	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	90	S52	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	91	IND8	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	92	IND5	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	93	SW2	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	94	HK2	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	95	HK3	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	96	HK4	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	97	HK5	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	98	HK6	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	99	HK7	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	100	HK8	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	101	HK9	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		
VEVRNtS8S-YA	102	HK10	S-n-lTHoLTHAVL-l	V-ERNGTLH	IOVT-NVAVRhgALThmL	AHL-HIVHVA-TV	AL-V	-v	AVMHV		

**FIGURE 3**

Genotype	Seq ID No:	Isolate	290	300	310	320	330	340	350	360	370	380
IV/2b	82	DK11	8-AFIVS-ERHHT-E	I-G-HIT	L-TLTHLAYAN-V-ELVLEVV-G	VVFGL-YFSHOG-A	IAILLVA-DA					
	83	SW3	8-AFIVS-ERHHT-E	I-G-HIT	L-TLTHLAYAN-V-ELVLEVV-G	VVFGL-YFSHOG-A	IAILLVA-DA					
	84	DK8	8-AFIVS-ERHHT-E	I-G-HIT	L-TLTHLAYAN-V-ELVLEVV-G	VVFGL-YFSHOG-A	IAILLVA-DA					
	85	SW8	8-AFIVS-ERHHT-E	I-G-HIT	L-TLTHLAYAN-V-ELVLEVV-G	VVFGL-YFSHOG-A	IAILLVA-DA					
22c	86	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
	87	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
	88	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
	89	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
III/2a	90	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
	91	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
	92	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
	93	SW3	A-VVVS-QHKEV-E	I-P-KIT	H-TATHLAYAN-V-EVLIDIV-G	VHFGI-YFSHOG-A	IVILLVA-DA					
(V)/3a	94	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
	95	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
	96	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
	97	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
II/1b	98	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
	99	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
	100	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
	101	SW3	G-AFTR-RRHVT-T	L-P-HLS	M-AVGHVAEVL-L-OTLFDIIA-A	IAAGL-YFSHOG-A	AIHMF8-DA					
I/1a	102	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	103	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	104	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	105	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
4a	106	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	107	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	108	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	109	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
4c	110	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	111	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	112	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	113	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
4d	114	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	115	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	116	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	117	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
4b	118	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	119	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	120	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	121	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
5a	122	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	123	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	124	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	125	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
6a	126	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	127	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	128	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					
	129	SW3	G-LFTS-RRHET-D	I-P-HVS	M-TAALVVSOLI-I-QAVDHVA-A	IAAGL-YFSHOG-A	IAILLVA-DA					

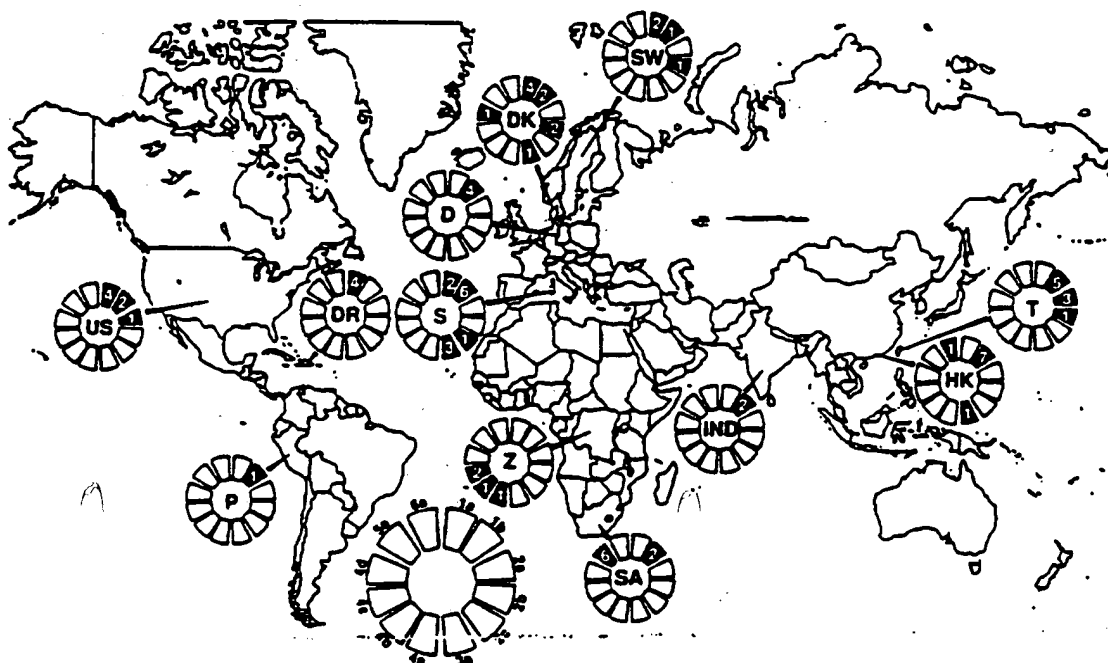
46/89

FIGURE 4



47/89

FIGURE 5



48/89

FIGURE 6A

SEQ ID NO: ISOLATE  
 108 DR4  
 103 DK7  
 104 US11  
 105 S14  
 106 SW1  
 107 S18

103-108 consensus

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCACAGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCACAGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCACAGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCACAGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCACAGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCACAGG

ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGTCGCCACAGG

SEQ ID NO: ISOLATE  
 108 DR4  
 103 DK7  
 104 US11  
 105 S14  
 106 SW1  
 107 S18

103-108 consensus

62 ACGTCAAGTTCCCGGGTGGCGGTGAGATCGTTGGTGGAGTTTACTTGTGTCGCGCAGGGG  
 62 ACGTCAAGTTCCCGGGTGGCGGTGAGATCGTTGGTGGAGTTTACTTGTGTCGCGCAGGGG  
 62 ACGTCAAGTTCCCGGGTGGCGGTGAGATCGTTGGTGGAGTTTACTTGTGTCGCGCAGGGG  
 62 ACGTCAAGTTCCCGGGTGGCGGTGAGATCGTTGGTGGAGTTTACTTGTGTCGCGCAGGGG  
 62 ACGTCAAGTTCCCGGGTGGCGGTGAGATCGTTGGTGGAGTTTACTTGTGTCGCGCAGGGG  
 62 ACGTCAAGTTCCCGGGTGGCGGTGAGATCGTTGGTGGAGTTTACTTGTGTCGCGCAGGGG

ACGTCAAGTTCCCGGGTGGCGGTGAGATCGTTGGTGGAGTTTACTTGTGTCGCGCAGGGG

SEQ ID NO: ISOLATE  
 108 DR4  
 103 DK7  
 104 US11  
 105 S14  
 106 SW1  
 107 S18

103-108 consensus

123 CCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGA  
 123 CCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGA  
 123 CCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGA  
 123 CCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGA  
 123 CCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGA  
 123 CCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGA

CCTAGATTGGGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTGCGCAACCTCGAGGTAGA

SEQ ID NO: ISOLATE  
 108 DR4  
 103 DK7  
 104 US11  
 105 S14  
 106 SW1  
 107 S18

103-108 consensus

184 CGTCAGCCTATCCCCAAGGCGCGTCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  
 184 CGTCAGCCTATCCCCAAGGCGCGTCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  
 184 CGTCAGCCTATCCCCAAGGCGCGTCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  
 184 CGTCAGCCTATCCCCAAGGCGCGTCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  
 184 CGTCAGCCTATCCCCAAGGCGCGTCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC  
 184 CGTCAGCCTATCCCCAAGGCGCGTCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC

CGTCAGCCTATCCCCAAGGCGCGTCCGGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC

SEQ ID NO: ISOLATE  
 108 DR4  
 103 DK7  
 104 US11  
 105 S14  
 106 SW1  
 107 S18

103-108 consensus

245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGG  
 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGG  
 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGG  
 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGG  
 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGG  
 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGG

CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCCCGTGG

SEQ ID NO: ISOLATE  
 108 DR4  
 103 DK7  
 104 US11  
 105 S14  
 106 SW1  
 107 S18

103-108 consensus

306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTC  
 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTC  
 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTC  
 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTC  
 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTC  
 306 CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTC

CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTGCGCAATTGGGTAAGGTC

49/89

FIGURE 6A

<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
108	DR4	367 ATCGAcACCCTcACGTGCGGCTTCGCCGACCTCATGGGGTACATcCCGCTCGTCGGCGCCC
103	DK7	367 ATCGATAACCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
104	US11	367 ATCGATAACCTTACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
105	S14	367 ATCGATAACCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATACCGCTCGTCGGCGCCC
106	SW1	367 ATCGATAACCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
107	S18	367 ATCGATAACCTCACGTGCGGCTTCGCCGACCTCATGGGGTACATTCCGCTCGTCGGCGCCC
103-108	consensus	ATCGAtACCCTcACGTGCGGCTTCGCCGACCTCATGGGGTACATaCCGCTCGTCGGCGCCC
<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
108	DR4	428 CcCTTGGgGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGaGTTCTGGAAGACGGCGTGAA
103	DK7	428 CTCTTGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
104	US11	428 CTCTCGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
105	S14	428 CcCTCGGgGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
106	SW1	428 CTCTtGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
107	S18	428 CTCTcGGAGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAA
103-108	consensus	CtCT- GGaGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGgGTTCTGGAAGACGGCGTGAA
<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
108	DR4	489 CTATGCAACAGGGAAtCTTCCTGGTTGCTCTTTCTCTATCTTCCTTTTGGCcttTGCTCTCT
103	DK7	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTTTGGCCCTGCTCTCT
104	US11	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCCCTGCTCTCT
105	S14	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTcCTaGCCCTGCTTTCT
106	SW1	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCCCTGCTTTCT
107	S18	489 CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCCCTGCTtCTCT
103-108	consensus	CTATGCAACAGGGAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTtctTgGCcctTGCTCTCT
<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
108	DR4	550 TGctTGACCGTGCCCGCaTCGGCC
103	DK7	550 TGCCTGACCGTGCCCGCTTCGGCC
104	US11	550 TGCCTGACTGTGCCCGCTTCAGCC
105	S14	550 TGCCTGACTGTGCCCGCTTCAGCC
106	SW1	550 TGCCTGACaGTGCCCGCGTCAGCC
107	S18	550 TGtCTGACTGTGCCCGCGTCAGCt
103-108	consensus	TGcctGACTGTGCCCGCctTCaGCC



50/89

**FIGURE 6B**

SEQ ID NO: ISOLATE

119 S9

117 IND3

118 IND8

111 D1

112 US6

113 P10

114 DK1

115 T10

116 SW2

122 HK4

109 SA10

110 S45

123 P8

124 T3

120 HK3

121 HK5

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

109-124 consensus

ATGAGCACGAATCCTAAACCTCAAAGAAAAACC

SEQ ID NO: ISOLATE

119 S9

117 IND3

118 IND8

111 D1

112 US6

113 P10

114 DK1

115 T10

116 SW2

122 HK4

109 SA10

110 S45

123 P8

124 T3

120 HK3

121 HK5

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

62 ACGTAAAGTTC

109-124 consensus

ACGTAAAGTTC

SEQ ID NO: ISOLATE

119 S9

117 IND3

118 IND8

111 D1

112 US6

113 P10

114 DK1

115 T10

116 SW2

122 HK4

109 SA10

110 S45

123 P8

124 T3

120 HK3

121 HK5

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

123 CCCCAGGTTGGGTGTGCGCGc

109-124 consensus

CCCCAGGTTGGGTGTGCGCGc

SEQ ID NO: ISOLATE

119 S9

117 IND3

118 IND8

111 D1

112 US6

113 P10

114 DK1

115 T10

116 SW2

122 HK4

109 SA10

110 S45

123 P8

124 T3

120 HK3

121 HK5

184 CGACAACCTATCCCCAAGGCTCGCCat

51/89

FIGURE 6B

117	IND3	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
118	IND8	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
111	D1	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
112	US6	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
113	P10	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
114	DK1	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
115	T10	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
116	SW2	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
122	HK4	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
109	SA10	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
110	S45	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
123	P8	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
124	T3	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
120	HK3	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
121	HK5	184	CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC

109-124 consensus

CGACAACCTATCCCCAAGGCTCGCCGGCCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC

SEQ ID NO: ISOLATE

119	S9	245	CTTGGCCCCCTCTACGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
117	IND3	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
118	IND8	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
111	D1	245	CTTGGCCCCCTCTATGGCAACGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
112	US6	245	CTTGGCCCCCTCTATGGCAACGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
113	P10	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
114	DK1	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
115	T10	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
116	SW2	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
122	HK4	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
109	SA10	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
110	S45	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
123	P8	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
124	T3	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
120	HK3	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG
121	HK5	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG

109-124 consensus

CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTACCCCGCGG

SEQ ID NO: ISOLATE

119	S9	306	CTCTCGGCCTAGTTGGGGCCCCAaGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
117	IND3	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
118	IND8	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
111	D1	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
112	US6	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
113	P10	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
114	DK1	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
115	T10	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
116	SW2	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
122	HK4	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
109	SA10	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
110	S45	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
123	P8	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
124	T3	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
120	HK3	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC
121	HK5	306	CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC

109-124 consensus

CTCTCGGCCTAGTTGGGGCCCCCAGACCCCCCGCGTAGGTCGCGTAATTGGGTAAGGTC

SEQ ID NO: ISOLATE

119	S9	367	ATCGATACCCTCACATGCGGCTTcGCCGACCTCATGGGGTACATcCGCTCGTCGGCGCCC
117	IND3	367	ATCGATACCCTCACATGCGGCTTcGCCGACCTCATGGGGTACATcCGCTCGTCGGCGCCC
118	IND8	367	ATCGATACCCTCACATGCGGCTTcGCCGACCTCATGGGGTACATcCGCTCGTCGGCGCCC

52/89

FIGURE 6B

111	D1	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
112	US6	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
113	P10	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
114	DK1	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
115	T10	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
116	SW2	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
122	HK4	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
109	SA10	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
110	S45	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGCCC
123	P8	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGGCC
124	T3	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGGCC
120	HK3	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGGCC
121	HK5	367	ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGGCC

109-124 consensus

ATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGGTACATCCCGCTCGTCGGCGGCC

SEQ ID NO: ISOLATE

119	S9	428	CCCTAGGGGGCGCTGCCAGGGCTCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
117	IND3	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
118	IND8	428	CCCTAGGGGGTGTCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
111	D1	428	CCCTAGGGGGTGTCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
112	US6	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
113	P10	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
114	DK1	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
115	T10	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
116	SW2	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
122	HK4	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
109	SA10	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
110	S45	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
123	P8	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
124	T3	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
120	HK3	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA
121	HK5	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA

109-124 consensus

CccTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTCTGGAGGACGGCGTGAA

SEQ ID NO: ISOLATE

119	S9	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTCTGCTTTGCTGTCC
117	IND3	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
118	IND8	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
111	D1	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
112	US6	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
113	P10	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
114	DK1	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
115	T10	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
116	SW2	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
122	HK4	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
109	SA10	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
110	S45	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
123	P8	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
124	T3	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
120	HK3	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC
121	HK5	489	CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTTGGCTTTGCTATCC

109-124 consensus

CTATGCAACAGGGAACCTGCCCGGTGCTCTTTCTCTATCTTCCCTTTGCTTTGCTATCC

SEQ ID NO: ISOLATE

119	S9	550	TGTTTGACCATCCAGCTTCCGCT
117	IND3	550	TGTTTGACCATCCAGCTTCCGCT
118	IND8	550	TGTTTGACCATCCAGCTTCCGCT
111	D1	550	TGTTTGACCATCCAGCTTCCGCT
112	US6	550	TGTTTGACCATCCAGCTTCCGCT

53/89

## FIGURE 6B

113	P10	550	TGccTGACCATCCCAGCgTCCGCT
114	DK1	550	TGTtTGACCATCCCAGCTTCCGCc
115	T10	550	TGTCTGACCATCCCAGCTTCCGCT
116	SW2	550	TGTCTGACCATCCCAGCTTCCGCT
122	HK4	550	TGTTTGACCATCCCAGCTTCCGCT
109	SA10	550	TGTTTgACCATCCCAGCTTCCGCT
110	S45	550	TGcTTGACCATCCCAGCTTCCGCT
123	P8	550	TGtcTGACCATCCCAGCTTCCGCT
124	T3	550	TGCTTGACCATCCCAGCTTCCGCT
120	HK3	550	TGCTTGACCACCCAGCTTCCGCT
121	HK5	550	TGtcTGACCACCCAGtTCCGCT
109-124	consensus		TGttTgACCatcCCAGctTCCGct

54/89

FIGURE 6C

SEQ ID NO:	ISOLATE	
119	S9	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
117	IND3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
118	IND8	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
111	D1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
112	US6	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
113	P10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
114	DK1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
115	T10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
116	SW2	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
122	HK4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
109	SA10	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
110	S45	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
123	P8	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
124	T3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
120	HK3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
121	HK5	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
108	DR4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
104	US11	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
105	S14	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
106	SW1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
107	S18	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG
103	DK7	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG

103-124 consensus

ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCGCCACAGG

SEQ ID NO:	ISOLATE	
119	S9	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTcGGTGGAGTTTACCTGTTGCCGCGCAGGGG
117	IND3	62 ACGTCAAGTTCCCGGGCGGTGGTGGCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
118	IND8	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
111	D1	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
112	US6	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
113	P10	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
114	DK1	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
115	T10	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
116	SW2	62 ACGTCAAGTTCCCGGGCGGTGGCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
122	HK4	62 ACGTCAAGTTCCCGGGCGGTGGCCAGATCGTcGGTGGAGTTTACCTGTTGCCGCGCAGGGG
109	SA10	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
110	S45	62 ACGTCAAGTTCCCGGGcGGcGGTcCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
123	P8	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
124	T3	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
120	HK3	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
121	HK5	62 ACGTCAAGTTCCCGGGCGGTGGTcCAGATCGTGGTGGAGTTTACCTGTTGCCGCGCAGGGG
108	DR4	62 ACGTCAAGTTCCCGGGTGGCGGTcCAGATCGTGGTGGAGTTTACTTGTGGCCGCGCAGGGG
104	US11	62 ACGTCAAGTTCCCGGGTGGCGGTcCAGATCGTGGTGGAGTTTACTTGTGGCCGCGCAGGGG
105	S14	62 ACGTCAAGTTCCCGGGTGGCGGTcCAGATCGTGGTGGAGTTTACTTGTGGCCGCGCAGGGG
106	SW1	62 ACGTCAAGTTCCCGGGTGGCGGTcCAGATCGTGGTGGAGTTTACTTGTGGCCGCGCAGGGG
107	S18	62 ACGTCAAGTTCCCGGGTGGCGGTcCAGATCGTGGTGGAGTTTACTTGTGGCCGCGCAGGGG
103	DK7	62 ACGTCAAGTTCCCGGGTGGCGGTcCAGATCGTGGTGGAGTTTACTTGTGGCCGCGCAGGGG

103-124 consensus

ACGTCAAGTTCCCGGGcGGcGGTcCAGATCGTcGGTGGAGTTTAcCTGTTGCCGCGCAGGGG

SEQ ID NO:	ISOLATE	
119	S9	123 CCCCAGGTTGGGTGTGCGCGCgACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
117	IND3	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
118	IND8	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
111	D1	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
112	US6	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
113	P10	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
114	DK1	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
115	T10	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
116	SW2	123 CCCCcGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG
122	HK4	123 CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTGCGAACCTCGTGGAAAGG

55/89

FIGURE 6C

109	SA10	123	CCCCAGGTTGGGTGTGCGCGCGAGcAGGAAGACTTCCGAGCGGTTCGCAACCTCGTGGAAAGG
110	S45	123	CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTcCAACCTCGTGGAAcGG
123	P8	123	CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGaTCGCAACCTCGTGGcAGG
124	T3	123	CCCCAGGTTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTTCGCAACCTCGTGGAAAGG
120	HK3	123	CCCCAGGTTGGGTGTGCGCGCGACcAGGAAGACTTcAGAGCGGTTCGCAACCTCGTGGAAAGG
121	HK5	123	CCCCAGGTTGGGTGTGCGCGCGAGCAGGAAGACTTCCGAGCGGTTCGCAACCTCGTGGAAAGG
108	DR4	123	CCCTAGATTGGGTGTGCGCGCGAGGAAAGACTTCCGAGCGGTTCGCAACCTCGAGGTAGA
104	US11	123	CCCTAGATTGGGTGTGCGCGCGAGGAAAGACTTCCGAGCGGTTCGCAACCTCGAGGTAGA
105	S14	123	CCCTAGATTGGGTGTGCGCGCGAGGAAAGACTTCCGAGCGGTTCGCAACCTCGAGGTAGA
106	SW1	123	CCCTAGATTGGGTGTGCGCGCGAGGAAAGACTTCCGAGCGGTTCGCAACCTCGAGGTAGA
107	S18	123	CCCTAGATTGGGTGTGCGCGCGAGGAAAGACTTCCGAGCGGTTCGCAACCTCGcGGTAGA
103	DK7	123	CCCTAGATTGGGTGTGCGCGCGcAGGAAGACTTCCGAGCGGTTCGCAACCTCGaGGTAGA

103-124 consensus CCCCAGgTTGGGTGTGCGCGCGaCtAGGAAGACTTCCGAGCGGTcGCAACCTCGtGGAAgG

SEO ID NO: ISOLATE

119	S9	184	CGACAACCTATCCCCAAGGCTCGCCatCCCCAGGGcAGGGCCTGGGCTCAGCCCGGGTACC
117	IND3	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
118	IND8	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGTAGGGCTGGGCTCAGCCCGGGcACC
111	D1	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
112	US6	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
113	P10	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
114	DK1	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
115	T10	184	CGACAgCCTATCCCCAAGGCTCGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCCGGGTACC
116	SW2	184	CGACAACCTATCCCCAAGGCTCGCCAGCCCGAGGGCAGGGCCTGGGCTCAGCCcGGGTACC
122	HK4	184	CGACAACCTATCCCCAAGGCTCGCCaCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
109	SA10	184	CGACAACCTATCCCCAAGGCTCGCCAGCCCGAGGGCAGGACCTGGGCCCAGCCCGGGTACC
110	S45	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGCAGGGCCTGGGCCCAGCCCGGGCaC
123	P8	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGTAGGGCTGGGCTCAGCCCGGGCACC
124	T3	184	CGACAACCTATCCCCAAGGCTCGCCGCGCCGAGGGTAGGGCTGGGCTCAGCCCGGGTACC
120	HK3	184	CGACAACCTATCCCCAAGGCTCGCCaACCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
121	HK5	184	CGACAACCTATCCCCAAGGCTCGCCGACCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
108	DR4	184	CGTCAGCCTATCCCCAAGGcCGTCGCGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
104	US11	184	CGTCAGCCTATCCCCAAGGCACGTGCGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
105	S14	184	CGTCAGCCTATCCCCAAGGCACGTGCGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
106	SW1	184	CGTCAGCCTATCCCCAAGGCGCGTCGCGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTATC
107	S18	184	CGTCAGCCTATCCCCAAGGCGCGTCGCGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC
103	DK7	184	CGTCAGCCTATCCCCAAGGCaCGTCGCGCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACC

103-124 consensus CGaCAaCCTATCCCCAAGGCTCGcCgGCCGAGGGcAGGgCCTGGGCTCAGCCcGGGTACC

SEO ID NO: ISOLATE

119	S9	245	CTTGGCCCCCTCTaCGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGtGG
117	IND3	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGCGG
118	IND8	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGCGG
111	D1	245	CTTGGCCCCCTCTATGGCAACGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGCGG
112	US6	245	CTTGGCCCCCTCTATGGCAACGAGGGCaTGGGGTGGGCAGGATGGCTCCTGTcACCCCGTGG
113	P10	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGTGG
114	DK1	245	CTTGGCCCCCTCTATGGCAATGAGGGCATGGGGTGGGCAGGATGGCTCCTGTcACCCCGcGG
115	T10	245	CTTGGCCCCCTCTATGGCAATGAGGGCATGGGGTGGGCAGGATGGCTCCTGTcACCCCGtGG
116	SW2	245	CcTGGCCCCCTCTATGGCAATGAGGGCATGGGaTGGGCAGGATGGCTCCTGTcCccCGCGG
122	HK4	245	CTTGGCCCCCTCTATGGCAATGAGGGCATGGGGTGGGCAGGATGGCTCCTGTcACCCCGCGG
109	SA10	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGTGG
110	S45	245	CTTGGCCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGTGG
123	P8	245	CTTGGCCCCCTCTATGcCAATGAGGGCTTGGGGTGGGCAGGATGGCTCCTGTcACCCCGCGG
124	T3	245	CTTGGCCCCCTCTATGGCaACGAGGGCATGGGGTGGGCAGGATGGCTCCTGTcACCCCGCGG
120	HK3	245	CTTGGCCCCCTCTATGGCAACGAGGGCATGGGGTGGGCAGGATGGCTCCTGTcACCCCGCGG
121	HK5	245	CTTGGCCCCCTCTATGGCAATGAGGGCATGGGGTGGGCAGGATGGCTCCTGTcACCCCaTGG
108	DR4	245	CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCAGGATGGCTCCTGTcCccCGTGG
104	US11	245	CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCAGGATGGCTCCTGTcTCCCCGTGG
105	S14	245	CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCAGGATGGCTCCTGTcTCCCCGTGG
106	SW1	245	CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGaTGGGCAGGATGGCTCCTGTcCCCCGTGG
107	S18	245	CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCAGGATGGCTCCTGTcCCCCGTGG

56/89

FIGURE 6C

103 DK7 245 CTTGGCCCCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCtCCCCGTGG  
 103-124 consensus CcTGGCCCCCTCTAtGgCaAtGAGGGCttgGGgTGGGCaGGATGGCTCCTGTCaCCCCgtGG

## SEQ ID NO: ISOLATE

119	S9	306	CTCTCGGCCTAGTTGGGGCCCCAaGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
117	IND3	306	tTCTCGGCCTAGTTGGGGCCCCACAGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
118	IND8	306	CTCTCGGCCTAGTTGGGGCCCCACAGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
111	D1	306	CTCCCGGCCTAGTTGGGGCCCCACaGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
112	US6	306	CTCCCGGCCTAGTTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
113	P10	306	CTCTCGGCCTAGTTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
114	DK1	306	CTCTCGGCCTAGTTGGGGCCCCAaGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
115	T10	306	CTCcCGGCCTAGTTGGGGCCCCACaGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
116	SW2	306	CTCTCGGCCTAGTTGGGGCCCCACtGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
122	HK4	306	CTCTCGGCCTAGTTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
109	SA10	306	CTCTCGGCCTAGTTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
110	S45	306	CTCCCGGCCTAGTTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
123	P8	306	CTCCCGGCCTAGTTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
124	T3	306	CTCCCGGCCTAATTGGGGCCCCACaGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
120	HK3	306	CTCTCGGCCTAATTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
121	HK5	306	CTCTCGGCCTAGTTGGGGCCCCACGGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
108	DR4	306	CTCTCGGCCTAGCTGGGGCCCCACaGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
104	US11	306	CTCTCGGCCTAGCTGGGGCCCCACgGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
105	S14	306	CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
106	SW1	306	CTCTCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
107	S18	306	CTCcCGGCCTAGCTGGGGCCCCACAGACCCCGGCGTAGGTCGCGTAATTGGGTAAGGTC
103	DK7	306	CTCtCGGCCTAGCTGGGGCCCCACAGACCCCGGCGcAGGTCGCGTAATTGGGTAAGGTC
103-124	consensus		cTctCGGCCTAGtTGGGGCCCCAc-GACCCCGGCGtAGGTCGCGTAATTGGGTAAGGTC

## SEQ ID NO: ISOLATE

119	S9	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
117	IND3	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
118	IND8	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
111	D1	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
112	US6	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
113	P10	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
114	DK1	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
115	T10	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
116	SW2	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
122	HK4	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
109	SA10	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
110	S45	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
123	P8	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
124	T3	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
120	HK3	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
121	HK5	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
108	DR4	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
104	US11	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
105	S14	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
106	SW1	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
107	S18	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
103	DK7	367	ATCGATACCCTCACATGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGCCC
103-124	consensus		ATCGATACCCTcACaTGCGGCTTtGCCGACCTCATGGGGTACATtCCGCTCGTCGGCGccc

## SEQ ID NO: ISOLATE

119	S9	428	CCCTAGGGGGCGCTGCCAGGGCctTGCGCGATGGCGTCCGGGTtCTGGAGGACGGCGTGAA
117	IND3	428	CCCTAGGGGGCGCTGCCAGGGCCTTGCGCGATGGCGTCCGGGTCTGGAGGACGGCGTGAA
118	IND8	428	CCCTAGGGGGCGCTGCCAGGGCCTTGCGCGATGGCGTCCGGGTCTGGAGGACGGCGTGAA
111	D1	428	CCCTAGGGGGTCTGCCAGGGCCTTGCGCGATGGCGTCCGGGTCTGGAGGACGGCGTGAA
112	US6	428	CCCTAGGGGGCGCTGCCAGGGCctTGCGCGATGGCGTCCGGGTCTGGAGGACGGCGTGAA

57/89

FIGURE 6C

113	P10	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
114	DK1	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
115	T10	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
116	SW2	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
122	HK4	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
109	SA10	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
110	S45	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
123	P8	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
124	T3	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
120	HK3	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
121	HK5	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
108	DR4	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
104	US11	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
105	S14	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
106	SW1	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
107	S18	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA
103	DK7	428	CCCTAGGGGGCGCTGCCAGGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAA

103-124 consensus

CccTaGGgGGcGcTGCCAGgGCcctGGCGcCatGGcGTCCGgGTtcTGGAgGACGGCGTGAA

SEQ ID NO: ISOLATE

119	S9	489	CTATGCAACAGGGAACctcCCCGGTTGCTCTTTCTCTATCTTCCTTctTgGCTTTGCTgTCC
117	IND3	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTaGCTTTGCTATCC
118	IND8	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
111	D1	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
112	US6	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
113	P10	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
114	DK1	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
115	T10	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
116	SW2	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
122	HK4	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
109	SA10	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
110	S45	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
123	P8	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
124	T3	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
120	HK3	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
121	HK5	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
108	DR4	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
104	US11	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
105	S14	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
106	SW1	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
107	S18	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC
103	DK7	489	CTATGCAACAGGGAACctTgCCCGGTTGCTCTTTCTCTATCTTCCTTTTGGCTTTGCTATCC

103-124 consensus

cTAtGCAACAGGGAACctTgCCCGGTTGCTCtTTcTCTATCTTCCTcTgGCTtTGcTgTCC

SEQ ID NO: ISOLATE

119	S9	550	TGTTTGACCATCCAGCTTCCGCT
117	IND3	550	TGTTTGACCATCCAGCTTCCGCT
118	IND8	550	TGTTTGACCATCCAGCTTCCGCT
111	D1	550	TGTTTGACCATCCAGCTTCCGCT
112	US6	550	TGTTTGACCATCCAGCTTCCGCT
113	P10	550	TGcCTGACCATCCAGCgTCCGCT
114	DK1	550	TGTTTGACCATCCAGCTTCCGCT
115	T10	550	TGTTTGACCATCCAGCTTCCGCT
116	SW2	550	TGTTTGACCATCCAGCTTCCGCT
122	HK4	550	TGTTTGACCATCCAGCTTCCGCT
109	SA10	550	TGTTTaACCATCCAGCTTCCGCT
110	S45	550	TGcTTGACCATCCAGCTTCCGCT
123	P8	550	TGcCTGACCATCCAGCTTCCGCT
124	T3	550	TGCTTGACCATCCAGCTTCCGCT
120	HK3	550	TGCTTGACCATCCAGCTTCCGCT
121	HK5	550	TGcCTGACCATCCAGCTTCCGCT



58/89

## FIGURE 6C

108	DR4	550	TGctTGACCGTGCCCGCaTCgGCC
104	US11	550	TGCCTGACTGTGCCCGCTTCAGCC
105	S14	550	TGCCTGACTGTGCCCGCTTCAGCC
106	SW1	550	TGCCTGACaGTGCCCGCGTCAGCC
107	S18	550	TGtCTGACTGTGCCCGCGTCAGct
103	DK7	550	TGcCTGACcGTGCCCGCctTCgGCC
103-124	consensus		TGtTgACcatcCCaGctTCcGct

59/89

FIGURE 5D

SEO ID NO: ISOLATE

128 T2  
125 T4  
126 US10  
127 T9

125-128 consensus

1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACACcAACCGTCGCCACAAg  
1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACACcAACCGTCGCCACAAg  
1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACACcAACCGTCGCCACAAg  
1 ATGAGCACAAATCCcAAACCCcCAAAGAAAAACCAcAAGAAACACcAACCGTCGCCACAAg

ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACAC - AACCGTCGCCACAA - G

SEO ID NO: ISOLATE

128 T2  
125 T4  
126 US10  
127 T9

125-128 consensus

62 ACGTTAAGTTcCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGcTGCCGCGCAGGGG  
62 ACGTTAAGTTcCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGGCGCGCAGGGG  
62 ACGTTAAGTTcCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGGCGCGCAGGGG  
62 ACGTTAAGTTcCCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGTTGGCGCGCAGGGG

ACGTTAAGTT - CCGGGCGGCGGCCAGATCGTTGGCGGAGTATACTTGcTGCCGCGCAGGGG

SEO ID NO: ISOLATE

128 T2  
125 T4  
126 US10  
127 T9

125-128 consensus

123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCGGAGCGgTCCAGCCcCGTGGaAGG  
123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCGGAGCGaTCCAGCCcCGTGGAGG  
123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCGGAGCGGTCCAGCCcCGTGGAGG  
123 CCCcAGGTTGGGTGTGCGCaCGACAAGGAAGACTTCGGAGCGGTCCAGCCcCGTGGAGG

CCCcAGGTTGGGTGTGCGCGcCGACAAGGAAGACTTCGGAGCGgTCCAGCCcCGTGGgAGG

SEO ID NO: ISOLATE

128 T2  
125 T4  
126 US10  
127 T9

125-128 consensus

184 CGCCAGCCCATCCcTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAaAACcAGGATAcC  
184 CGCCAGCCCATCCcTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAaAACcAGGATAcC  
184 CGCCAGCCCATCCcTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAaAACcAGGATAcC  
184 CGCCAGCCCATCCcTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAaAACcAGGATAcC

CGCCAGCCCATCCcTAAAGATCGGCGCTCCACTGGCAAGTCCTGGGGAaAACcAGGATAcC

SEO ID NO: ISOLATE

128 T2  
125 T4  
126 US10  
127 T9

125-128 consensus

245 CCTGGCCCTGTATGGGAATGAGGGgCTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG  
245 CCTGGCCCTGTATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG  
245 CtTGGCCCTTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG  
245 CtTGGCCcCTATATGGGAATGAGGGACTCGGCTGGGCgGGATGGCTCCTGTCCCCCGAGG

CctTGGCCcCT - TATGGGAATGAGGGaCTCGGCTGGGCaGGATGGCTCCTGTCCCCCGAGG

SEO ID NO: ISOLATE

128 T2  
125 T4  
126 US10  
127 T9

125-128 consensus

306 TTCcCGTCCCTCtTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAAcGTGGGTAAgGTC  
306 TTCcCGTCCCTCtTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAAcGTGGGTAAgGTC  
306 TTCcCGTCCCTCtTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAAcGTGGGTAAgGTC  
306 TTCcCGTCCCTCtTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAAcGTGGGTAAgGTC

TTCcCGTCCCTCtTGGGGCCCCAATGACCCCCGGCATAGGTCGCGCAAcGTGGGTAAgGTC

SEO ID NO: ISOLATE

128 T2  
125 T4  
126 US10  
127 T9

125-128 consensus

367 ATCGATACCCTAACGTGCGgGCTTTGCCGACCTCATGGGGTACaTCCCCGTCTAGGCGcCC  
367 ATCGATACCCTAACGTGCGcGCTTTGCCGACCTCATGGGGTACgTCCCCGTCTAGGCGgCC  
367 ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGgTACATCCCCGTCTgGGCGcTC  
367 ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGgTACATCCCCGTCTgGGCGcCC

ATCGATACCCTAACGTGCGgGCTTTGCCGACCTCATGGGgTACaTCCCCGTCTgAGGCGcCC

SEO ID NO: ISOLATE

128 T2

428 CGcTtGGTGGtGTGCCAGAGCTCTtGCGCATGGCGTGAGAGTCCTGGAGGACGGaGTTAA

60/89

FIGURE 6D

125	T4	428	CGtTgGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGACGGGGTTAA
126	US10	428	CGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGgGTCCTGGAGGACGGGGTTAA
127	T9	428	CGCTTGGTGGCGTtGCCAGAGCTCTCGCGCAGGGCGTGAGaGTCCTGGAGGACGGGGTTAA
125-128	consensus		CGCTtGGTGGcGTcGCCAGAGCTCTcGGCGCatGGCGTGAGaGTCCTGGAGGACGGgGTTAA

<u>SEQ ID NO:</u> <u>ISOLATE</u>			
128	T2	489	TTATGCAACAGGtAACTTACCcGGTTGCTCCTTTTCTATcTTCTTGCTaGCCCTgCTGTCC
125	T4	489	TTATGCAACAGGGAACTTACCtGGTTGCTCCTTTTCTATtTTCTTGCTGGCCCTACTGTCC
126	US10	489	TTATGCAACAGGGAACTTACCcGGTTGCTCCTTTTCTATcTTCTTGCTGGCCtTACTGTCC
127	T9	489	TTATGCAACAGGGAAcTACCtGGTTGCTCtTTTCTATCtTTCTTGCTGGCCcTACTGTCC
125-128	consensus		TTATGCAACAGGgAAcTACC-TGGTTGCTCtTTTCTATcTTCTTGCTgGCCcTactGTCC

<u>SEQ ID NO:</u> <u>ISOLATE</u>			
128	T2	550	TGCATCACtATTCCgGtTTCaGCT
125	T4	550	TGCATCACCAATTCCAGTCTCcGCT
126	US10	550	TGCATCACCAATTCCAGTCTCTGCT
127	T9	550	TGCATCACCAcTCCgGcCTCTGCT
125-128	consensus		TGCATCACcAtTCC-GtctCTcGCT

61/89

FIGURE 6E

SEQ ID NO:	ISOLATE	
131	DK11	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
132	SW3	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
133	DK8	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACACAAACCGCGCCACAGG
129	T8	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACACAAACCGCGCCACAGG
130	US1	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACACAAACCGCGCCACAGG
129-133	consensus	ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAACACAAACCGCGCCACAGG
SEQ ID NO:	ISOLATE	
131	DK11	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
132	SW3	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
133	DK8	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
129	T8	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
130	US1	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
129-133	consensus	ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGCTGCCGCGCAGGGG
SEQ ID NO:	ISOLATE	
131	DK11	123 CCCCAGGTTGGGTGTGCGCGACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCTGGGAGA
132	SW3	123 CCCCAGGTTGGGTGTGCGCGACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCTGGGAGA
133	DK8	123 CCCCAGGTTGGGTGTGCGCGACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCTGGGAGA
129	T8	123 CCCCAGGTTGGGTGTGCGCGACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCTGGGAGA
130	US1	123 CCCCAGGTTGGGTGTGCGCGACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCTGGGAGA
129-133	consensus	CCCCAGGTTGGGTGTGCGCGACGACAAGGAAGACTTCCGAGCGATCCCAGCCGCTGGGAGA
SEQ ID NO:	ISOLATE	
131	DK11	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGcCCTGGGGAAAGCCAGGATATC
132	SW3	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCTTGGGGAAAGCCAGGATATC
133	DK8	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCTTGGGGAAAGCCAGGATATC
129	T8	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCTTGGGGAAAGCCAGGATATC
130	US1	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGTCTTGGGGAAAGCCAGGATATC
129-133	consensus	CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGcCCTGGGGAAAGCCAGGATATC
SEQ ID NO:	ISOLATE	
131	DK11	245 CTTGGCCCCCTGTATGGAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
132	SW3	245 CTTGGCCCCCTGTATGGAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
133	DK8	245 CTTGGCCCCCTGTATGGAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
129	T8	245 CTTGGCCCCCTGTATGGAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
130	US1	245 CTTGGCCCCCTGTATGGAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
129-133	consensus	CTTGGCCCCCTGTATGGAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
SEQ ID NO:	ISOLATE	
131	DK11	306 GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATAAATCACGCAATTGGGCAAAGTC
132	SW3	306 GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATAGATCACGCAATTGGGCAAAGTC
133	DK8	306 GTCTCGTCTACTTGGGGCCCCACTGACCCCCGGCATAGATCACGCAATTGGGCAAAGTC
129	T8	306 GTCTCGTCTACTTGGGGCCCCACTGACCCCCGGCATAGATCACGTAATTGGGCAAGTC
130	US1	306 GTCTCGTCTACTTGGGGCCCCACTGACCCCCGGCAcAGATCACGTAATTGGGCAAGTC
129-133	consensus	GTCTCgTCCTAcTTGGGGCCCCACTGACCCCCGGCAcAgATCACGcAAcTTGGGCAAAgTC
SEQ ID NO:	ISOLATE	
131	DK11	367 ATCGACACCATTACGTGTGGTTTTGCGGACCTCATGGGGTACATCCCTGTCTGTcGGCGCCC
132	SW3	367 ATCGACACCATTACGTGTGGTTTTGCGGACCTCATGGGGTACATCCCTGTCTGTcGGCGCCC
133	DK8	367 ATCGACACCATTACGTGTGGTTTTGCGGACCTCATGGGGTACATCCCTGTCTGTcGGCGCCC
129	T8	367 ATCGATACCATTACATGTGGTTTTGCGGACCTCATGGGGTACATCCCTGTCTGTcGGCGCCC

62/89

FIGURE 6E

130	US1	367	ATCGATACCATTACgTGTGGTTTTTGCCGACCTCATGGGGTACATCCCTGTCGTTGGCGCCC
129-133	consensus		ATCGAcACCATTACgTGTGGTTTTTGCCGACCTCATGGGGTACATCCCTGTCGtGGCGCCC
<u>SEQ ID NO:</u>	<u>ISOLATE</u>		
131	DK11	428	CGGTCGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTAGAGTCCTGGAAGACGGGATAAA
132	SW3	428	CGGTCGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTAGAGTCCTGGAAGACGGGATAAA
133	DK8	428	CGGTcGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTAGGGTCCTGGAAGACGGGATAAA
129	T8	428	CGGTCGGAGGCGTCGCCAGAGCTCTGGCACAcGGTGTAGGGTCCTGGAAGACGGGATAAA
130	US1	428	CGGTCGGAGGCGTCGCCAGAGCTCTGGCACAcGGTGTAGGGTCCTGGAAGACGGGATAAA
129-133	consensus		CGGTcGGAGGCGTCGCCAGAGCTCTGGCACAcGGTGTAGgGTCCTGGAAGACGGGATAAA
<u>SEQ ID NO:</u>	<u>ISOLATE</u>		
131	DK11	489	TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTCTATCTTCTTACTTGCTCTTCTGTCa
132	SW3	489	TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTCTATCTTCTTACTTGCTCTTCTGTcG
133	DK8	489	TTACGCAACAGGGAATTTGCCTGGTTGCTCTTTTCTATCTTCTTGTCTTGCTCTTCTGTcG
129	T8	489	cTAcGCAACAGGGAATTTGCCTGGTTGCTCTTTTCTATCTTCTTGTCTTGCTCTTCTGTcCa
130	US1	489	cTAcGCAACAGGGAATcTGCCTGGTTGCTCcTTTCTATCTTCTTAcTGTCTCTTCTGTcG
129-133	consensus		cTAcGCAACAGGGAATcTGCCTGGTTGCTCcTTTCTATCTTCTTAcTGTCTCTTCTGTcG
<u>SEQ ID NO:</u>	<u>ISOLATE</u>		
131	DK11	550	TGCTgCACAGTGCCAGTGTCTGCG
132	SW3	550	TGCTtCACAGTGCCAGTGTCTGCG
133	DK8	550	TGCTgCACAGTGCCAGTGTCTGCG
129	T8	550	TGCTtCACAGTGCCAGTGTCTGCA
130	US1	550	TGCgcCACgGTGCCgGTGTCTGCA
129-133	consensus		TGCT - CACaGTGCCaGTGTCTGCG

63/89

FIGURE 6F

SEQ ID NO:	ISOLATE	
131	DK11	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
132	SW3	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
133	DK8	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
129	T8C	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
130	US1	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
125	T4	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
126	US10	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
127	T9	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
128	T2	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
134	S83	1 ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
125-134	consensus	ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAGAAATACAAACCGCGCCACAGG
131	DK11	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
132	SW3	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
133	DK8	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
129	T8	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
130	US1	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
125	T4	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
126	US10	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
127	T9	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
128	T2	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
134	S83	62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
125-134	consensus	ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTACTTGCTGCCGCGCAGGGG
131	DK11	123 CCCAGGTTGGGTGTGCGCaCGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
132	SW3	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
133	DK8	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
129	T8	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
130	US1	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
125	T4	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
126	US10	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
127	T9	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
128	T2	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
134	S83	123 CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
125-134	consensus	CCCAGGTTGGGTGTGCGCGCAGACAAGGAAGACTTCCGAGCGATCCCAGCCGCGTGGGAGA
131	DK11	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
132	SW3	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
133	DK8	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
129	T8	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
130	US1	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
125	T4	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
126	US10	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
127	T9	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
128	T2	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
134	S83	184 CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
125-134	consensus	CGCCAGCCCATCCCGAAAGATCGGCGCTCCACCGGCAAGCTCTGGGGAAGCCAGGATATC
131	DK11	245 CTTGGCCCCCTGTATGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
132	SW3	245 CTTGGCCCCCTGTATGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
133	DK8	245 CTTGGCCCCCTGTATGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
129	T8	245 CTTGGCCCCCTGTATGGAAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG

64/89

FIGURE 6F

130	US1	245	CTTGGCCCTCTGTACGGAACGAGGGCTGCGGCTGGGCAGGTTGGCTCCTGTCCCCCGCGG
125	T4	245	CcTGGCCCTCTGTATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG
126	US10	245	CcTGGCCCTCTATATGGGAATGAGGGACTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG
127	T9	245	CCTGGCCCTCTATATGGGAATGAGGGACTCGGCTGGGCcGGATGGCTCCTGTCCCCCGAGG
128	T2	245	CCTGGCCCTGTATGGGAATGAGGGcCTCGGCTGGGCAGGATGGCTCCTGTCCCCCGAGG
134	S83	245	CcTGGCCCTGTATGGGAATGAGGGcCTCGGCTGGGCAGGgTGGCTCCTGTCCCCCGcGG
125-134	consensus		CcTGGCCcCTgTAtGG-AA-GAGGGc--CGGcTGGGCaGGcTGGCTCCTGTCCCCCGcGG
SEQ ID NO:	ISOLATE		
131	DK11	306	GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATAaATCAGCAATTGGGcAAAGTC
132	SW3	306	GTCTCATCCTAATTGGGGCCCCACTGACCCCCGGCATAGATCAGCAATTGGGGCAAAGTC
133	DK8	306	GTCTCGTCTACTTTGGGGCCCCACTGACCCCCGGCATAGATCAGCAATTGGGGCAAAGTC
129	T8	306	GTCTCGTCTACTTTGGGGCCCCACTGACCCCCGGCATAGATCAGCTAATTGGGGCAgAGTC
130	US1	306	GTCTCGTCTACTTTGGGGCCCCACTGACCCCCGGCaAGATCAGTAACTTTGGGCAAGGTC
125	T4	306	TTCCCGTCCCTCCTGGGGCCCCAaTGACCCCCGGCATAGGTTCGCGCAACGTGGGTAAAGTC
126	US10	306	TTCCCGTCCCTCCTGGGGCCCCAcTGAtCCCCGGCATAGGTTCGCGCAACGTGGGTAAAGTC
127	T9	306	TTCCCGTCCCTCCTGGGGCCCCAgTGACCCCCGGCATAGGTTCGCGCAACGTGGGTAAAGTC
128	T2	306	TTCTCGTCCCTCCTGGGGCCCCAaTGACCCCCGGCATAGGTTCGCGCAaTGGGTAAaGTC
134	S83	306	TTCTCGcCCcTCaTGGGGCCCCAcGACCCCCGGCATAaATCGCGCAAcTGGGTAAgGTC
125-134	consensus		-TCTCgtCCcT-cTGGGGCCCCAcTGAcCCCCGGCaTAgATC-CGcAA-cTGGGTAA-GTC
SEQ ID NO:	ISOLATE		
131	DK11	367	ATCGACACCATTACGTGTGGTTTTCGCCGACCTCATGGGGTACATCCCTGTCTGCGCGCCC
132	SW3	367	ATCGACACCATTACGTGTGGTTTTCGCCGACCTCATGGGGTACATCCCTGTCTGCGCGCCC
133	DK8	367	ATCGACACCATTACGTGTGGTTTTCGCCGACCTCATGGGGTACATCCCTGTCTGCGCGCCC
129	T8	367	ATCGATACCATTACaGTGTGGTTTTCGCCGACCTCATGGGGTACATCCCTGTCTGCGCGCCC
130	US1	367	ATCGATACCATTACGTGTGGTTTTCGCCGACCTCATGGGGTACATCCCTGTCTGCGCGCCC
125	T4	367	ATCGATACCCTAACGTGcAGCcTTGCCGACCTCATGGGGTACgTCCCGTCTGaGGCGcCC
126	US10	367	ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGaTACATCCCGTCTGgGGCGcTC
127	T9	367	ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCGTCTGAGGCGCCC
128	T2	367	ATCGATACCCTAACGTGCGGCTTTGCCGACCTCATGGGGTACATCCCGTCTGAGGCGCCC
134	S83	367	ATCGATACCCTAACGTGCGGcTTTGCCGACCTCATGGGGTACATaCCCGTCTGTGGCGcTC
125-134	consensus		ATCGATACC-T-ACgTG-gGtTTGCCGACCTCATGGGgTACATcCC-GTCGTGGCGccc
SEQ ID NO:	ISOLATE		
131	DK11	428	CGGTCCGAGGCGTCGCCAGAGCTCTGGCACACGGTGTAGAGTCTTGAAGACGGGATAAA
132	SW3	428	CGGTCCGAGGCGTCGCCAGAGCTCTGGCACACGGTGTAGAGTCTTGAAGACGGGATAAA
133	DK8	428	CGGTcGGAGGCGTCGCCAGAGCTCTGGCACACGGTGTAGGGTCTTGAAGACGGGATAAA
129	T8	428	CGGTCCGAGGCGTCGCCAGAGCTCTGGCACAcGGTGTAGGGTCTTGAAGACGGGATAAA
130	US1	428	CGGTCCGAGGCGTCGCCAGAGCTCTGGCACAcGGTGTAGGGTCTTGAAGACGGGATAAA
125	T4	428	CGcTgGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGaGTCCTGGAGGACGGGGTTAA
126	US10	428	CGCTTGGTGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGgGTCCTGGAGGACGGGGTTAA
127	T9	428	CGCTTGGTGGCGTcGCCAGAGCTCTCGCGCaGGCGTGAGAGTCTTGGAGGACGGGGTTAA
128	T2	428	CGCTTGGTGGtGTcGCCAGAGCTCTcGCGCATGGCGTGAGAGTCTTGGAGGACGgaGTTAA
134	S83	428	CcgTGGcGGcGTtGCCAGAGCcCTcGCGcCATGGgGTGAGgGTcCTGGAGGACGGGaTaAA
125-134	consensus		CggTcGGaGGcGTcGCCAGAGCcCTgGCaCA-GGtGT-AG-GTcCTGGA-GACGGGaTaAA
SEQ ID NO:	ISOLATE		
131	DK11	489	TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTCTATCTTCTTACTTGCTCTTCTGTCa
132	SW3	489	TTACGCAACAGGGAATCTGCCTGGTTGCTCTTTTCTATCTTCTTACTTGCTCTTCTGTcG
133	DK8	489	TTACGCAACAGGGAATTTGCCTGGTTGCTCTTTTCTATCTTCTTGTCTCTTCTGTcG
129	T8	489	cTAcGCAACAGGGAATTTGCCTGGTTGCTCTTTTCTATCTTCTTGTCTCTTCTGTcCa
130	US1	489	TTAcGCAACAGGGAATcTGcCTGGTTGCTCTTTTCTATCTTCTTaCTTGCTCTTCTGTcG
125	T4	489	TTATGCAACAGGGAACCTTACCTGGTTGCTCTTTTCTATcTTCTTGCTGGCCCTACTGTCC
126	US10	489	TTATGCAACAGGGAACCTTACCcGGTTGCTCTTTTCTATCTTCTTGCTGGCCcTACTGTCC
127	T9	489	TTATGCAACAGGGAACcTACCcGGTTGCTCTcTTTCTATCTTCTTGCTGGCCCTACTGTCC
128	T2	489	TTATGCAACAGGcAACTTACCCGGTTGCTCTTTTCTATCTTCTTGCTGACCTCTGTGCC

65/89

FIGURE 6F

134	S83	489	TTATGCAACgGGgAAAtTTgCCCGGTGCTCtTTtCTATCTTtTctTgGCCCTctTGTCt
125-134	consensus		tTatGCAACaGGgAAAtTgCCtGGTGTCTCtTTtCTATcTTctTgcTtGC-cTtcTGTCc
<u>SEQ ID NO:</u>	<u>ISOLATE</u>		
131	DK11	550	TGCTgCACAGTGCCAGTGTCTGCG
132	SW3	550	TGCTtCACAGTGCCAGTGTCTGCG
133	DK8	550	TGCTgCACAGTGCCAGTGTCTGCG
129	T8	550	TGCTtCACAGTGCCAGTGTCTGCA
130	US1	550	TGCgcCACgGTGCCgGTGTCTGCA
125	T4	550	TGCATCACCATTCCAGTCTCcGCT
126	US10	550	TGCATCACCATTCCAGTCTCTGCT
127	T9	550	TGCATCACCActCCGGcCTCTGCT
128	T2	550	TGCATCACTATTCCGGTTTCaGCT
134	S83	550	TGCATCtCTgTgCCaGTTTCcGCc
125-134	consensus		TGCatCaCagtGCCaGtgTCtGCT



66/89

FIGURE 6G

SEQ ID NO: ISOLATE

138 DK12  
135 HK10  
136 S52  
137 S2

135-138 consensus

1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAGAAACACCATCCGTCGCCACAGG  
1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAGAAACACCATCCGTCGCCACAGG  
1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAGAAACACCATCCGTCGCCACAGG  
1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAGAAACACCATCCGTCGCCACAGG

ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAAGAAACACCATCCGTCGCCACAGG

SEQ ID NO: ISOLATE

138 DK12  
135 HK10  
136 S52  
137 S2

135-138 consensus

62 ACGTCAAGTTCCTCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG  
62 ACGTTAAGTTCCTCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG  
62 ACGTCAAGTTCCTCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG  
62 ACATCAAGTTCCTCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG

ACGT- AAGTTCCTCCGGGTGGCGGACAGATCGTTGGTGGAGTATACGTGTTGCCGCGCAGGGG

SEQ ID NO: ISOLATE

138 DK12  
135 HK10  
136 S52  
137 S2

135-138 consensus

123 CCCACGATTGGGTGTGCGCGCAGCGCTAAACTTCTGAACGGTCaCAGCCTCGCGGACGg  
123 CCCACGATTGGGTGTGCGCGCAGCGCTAAACTTCTGAACGGTCgCAGCCTCGCGGACGA  
123 CCCACGATTGGGTGTGCGCGCAGCGCTAAACTTCTGAACGGTCACAGCCTCGCGGACGA  
123 CCCACGATTGGGTGTGCGCGCAGCGCTAAACTTCTGAACGGTCACAGCCTCGCGGACGg

CCCACGATTGGGTGTGCGCGCAGCGCTAAACTTCTGAACGGTCaCAGCCTCGCGGACG-

SEQ ID NO: ISOLATE

138 DK12  
135 HK10  
136 S52  
137 S2

135-138 consensus

184 CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGGTCTGGGCTCAGCCTGGGTACC  
184 CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGGTCTGGGCTCAGCCCCGGTACC  
184 CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGGTCTGGGCTCAGCCCCGGTACC  
184 CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGaTCCTGGGCTCAGCCCCGGTACC

CGACAGCCTATCCCCAAGGCGCGTCGGAGCGAAGGCCGgTCCTGGGCTCAGCCcGGGTACC

SEQ ID NO: ISOLATE

138 DK12  
135 HK10  
136 S52  
137 S2

135-138 consensus

245 CTTGGCCCCCTCTATGGTAACGAGGGCTGCGGGTGGGCAGGgTGGCTCCTGTCCCCACGCGG  
245 CTTGGCCCCCTCTATGGTAACGAGGGCTGCGGGTGGGCAGGaTGGCTCCTGTCCCCACGCGG  
245 CTTGGCCCCCTCTATGGTAACGAGGGCTGCGGGTGGGCAGGGTGGCTCCTGTCCCCACGCGG  
245 CTTGGCCCCCTCTATGGTAACGAGGGCTGCGGGTGGGCAGGGTGGCTCCTGTCCCCACGCGG

CTTGGCCCCCTCTATGGTAACGAGGGCTGCGGGTGGGCAGGgTGGCTCCTGTCCCCACGCGG

SEQ ID NO: ISOLATE

138 DK12  
135 HK10  
136 S52  
137 S2

135-138 consensus

306 CTCCCGTCCATCTTGGGGCCCAAACGACCCCCGGCGgaGGTCCCGCAATTGGGTAAAGTC  
306 CTCCCGTCCATCTTGGGGCCCAAACGACCCCCGGCGaCGGTCCCGCAATTGGGTAAAGTC  
306 CTCCCGTCCATCTTGGGGCCCAAACGACCCCCGGCGGAGGTCCCGCAATTGGGTAAAGTC  
306 CTCCCGTCCATCTTGGGGCCCAAAtGACCCCCGGCGGAGGTCCCGCAATTGGGTAAAGTC

CTCCCGTCCATCTTGGGGCCCAAACGACCCCCGGCGgaGGTCCCGCAATTGGGTAAAGTC

SEQ ID NO: ISOLATE

138 DK12  
135 HK10  
136 S52  
137 S2

135-138 consensus

367 ATCGATACCCTcACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC  
367 ATCGATACCCTTACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC  
367 ATCGATACCCTTACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC  
367 ATCGATACCCTTACGTGCGGcTTCGCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC

ATCGATACCCTcACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTCGGCGCTC

SEQ ID NO: ISOLATE

138 DK12

428 CcGTAGGgGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA

67/89

## FIGURE 6G

135	HK10	428	CCGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
136	S52	428	CCGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
137	S2	428	CCGTAGGAGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
135-138 consensus			CcGTAGGaGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA
<u>SEQ ID NO: ISOLATE</u>			
138	DK12	489	TTTCGCAACAGGGAAC TTGCCCGGTTGCTCCTTTTCTATCTTCCTTCCTGCTCTGTTCTCT
135	HK10	489	TTTCGCAACAGGGAAC TTGCCCGGTTGCTCCTTTTCTATCTTCCTTCCTGCTCTGTTCTCT
136	S52	489	TTTTCGCAACAGGGAAC TTGCCCGGTTGCTCCTTTTCTATCTTCCTTCCTGCTCTGTTCTCT
137	S2	489	TTTTCGCAACAGGGAAC TTGCCCGGTTGCTCCTTTTCTATCTTCCTTCCTGCTCTGTTCTCT
135-138 consensus			TTT-GCAACAGGGAAC TTGCCCGGTTGCTCCTTTTCTATCTTCCTTCCTGCTCTGTTCTCT
<u>SEQ ID NO: ISOLATE</u>			
138	DK12	550	TGCcTAATTCATCCAGCAGCTAGT
135	HK10	550	TGCTTAATTCATCCAGCAGCTAGT
136	S52	550	TGCTTAgtTCATCCcGCAGCTAGT
137	S2	550	TGCTTAaTTTCATCCaGCAGCTAGT
135-138 consensus			TGCcTAaTTTCATCCaGCAGCTAGT

68/89

FIGURE 6H

SEQ ID NO: ISOLATE  
 145 DK13  
 143 Z6  
 144 Z7  
 140 Z8  
 139 Z4  
 142 Z5  
 141 Z1

139-145 consensus

1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCATGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCATGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCATGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCATGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCATGG  
 1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCATGG

ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCATGG

SEQ ID NO: ISOLATE  
 145 DK13  
 143 Z6  
 144 Z7  
 140 Z8  
 139 Z4  
 142 Z5  
 141 Z1

139-145 consensus

62 ACGTTAAGTTCCCGGGTGGCGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG  
 62 ACGTTAAGTTCCCGGGTGGTGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG  
 62 ACGTTAAGTTCCCGGGCGGTGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG  
 62 AcGTAAATTTCCCGGGCGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG  
 62 ACGTAAATTTCCCGGGTGGTGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG  
 62 ATGTAAATTTCCCGGGTGGTGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG  
 62 ATGTgAAATTTCCCGGGCGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG

ACGT-AAATTTCCCGGGTGGTGGCCAGATCGTTGGCGGAGTTTACTTGTTCGCCGCGCAGGGG

SEQ ID NO: ISOLATE  
 145 DK13  
 143 Z6  
 144 Z7  
 140 Z8  
 139 Z4  
 142 Z5  
 141 Z1

139-145 consensus

123 CCCcAGaTTGGGTGTGCGCGCACTAGGAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg  
 123 CCCcAGgTTGGGTGTGCGCGCACTAGGAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg  
 123 CCCcAGaTTGGGTGTGCGCaCaACTAGGAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg  
 123 CCCcAGgTTGGGTGTGCGCGCACTCGGAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg  
 123 CCCcAGgTTGGGTGTGCGCGCACTCGaAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg  
 123 CCCcAGgTTGGGTGTGCGCGCACTCGGAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg  
 123 CCCcGGTGGGTGTGCGCGCaGCTCGGAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg

CCCcAGgTTGGGTGTGCGCGCaGCTCGGAAGACTTCGGAGCGGTTCGAACCTCGTGGGAGg

SEQ ID NO: ISOLATE  
 145 DK13  
 143 Z6  
 144 Z7  
 140 Z8  
 139 Z4  
 142 Z5  
 141 Z1

139-145 consensus

184 CGCCAGCCTATCCCCAAGGCgCGcCaActcGAGGGcAGGTCTGGGCTCAGCCtGGGTATC  
 184 CGCCAGCCTATCCCCAAGGCAGCTCGATCTGAGGGAAGGTCTGGGCTCAGCCCGGGTATC  
 184 CGTCAGCCTATCCCCAAGGCAGCTCGATCTGAGGGAAGGTCTGGGCTCaACCCGGGTACC  
 184 CGTCAGCCTATCCCCAAGGCAGCTCGGTcGAGGGcAGGTCTGGGCTCAGCCCGGGTACC  
 184 CGTCaACCTATCCCCAAGGCgCGcCaGcCaGAGGGCAGaTCTGGGcGcCAGCCCGGGTACC  
 184 CGTCAGCCTATCCCCcAGGCaCGtCGGTCCGAGGGCAGGTCTGGGCTCAGCCCGGGTACC  
 184 CGTCAGCCTATCCCCAAGGCgCGcCGGTCCGAGGGCAGGTCTGGGCTCAGCCCGGGTACC

CGTCAGCCTATCCCCAAGGCAGCTCGgtccGAGGGcAGGTCTGGGCTCAGCCCGGGTACC

SEQ ID NO: ISOLATE  
 145 DK13  
 143 Z6  
 144 Z7  
 140 Z8  
 139 Z4  
 142 Z5  
 141 Z1

139-145 consensus

245 CtTGGCCcCTTTACGGcAATGAGGGcTGCGGGTGGGCGGGATGGCTCCTGTcACCCCGTGG  
 245 CATGGCCCTCTTTACGGTAATGAGGGTTCGCGGTGGGCGGGATGGCTCCTGTcACCCCGTGG  
 245 CATGGCCCTCTTTACGGTAAGAGGGTTCGCGGTGGGCGGGATGGCTCCTGTcACCCCGTGG  
 245 CATGGCCCTCTTTACGGTAATGAAGGCTGTGGGTGGGcAGGcTGGCTCCTGTcACCCCGG  
 245 CTGGCCcCTcTATGGCAATGAGGGCTGcGGGTGGGcAGGGTGGCTCCTGTcCTcCGCGG  
 245 CTGGCCcCTTTATGGCAATGAGGGCTGTGGGTGGGcAGGGTGGCTCCTGTcCCCCCGGG  
 245 CTGGCCcCTTTACGGCAATGAGGGCTGTGGGTGGGcAGGGTGGCTCCTGTcCCCCCGGG

CtTGGCCcCTTTACGGcAATGAGGGcTGcGGGTGGGcAGG-TGGCTCCTGTc-CCcCGcGG

SEQ ID NO: ISOLATE  
 145 DK13  
 143 Z6  
 144 Z7  
 140 Z8

306 CTCCTCGgCCGTCTTGGGGcCCgAATGATCCCCGGCGgAGGTCCCGCAACTTGGGTAAAGTC  
 306 CTCCTCGACCGCTTGGGGcCCAAATGATCCCCGGCGAAGGTCCCGCAACTTGGGTAAAGTC  
 306 CTCCTCGACCGCTTGGGGcCCAAATGATCCCCGGCGAAGGTCCCGCAACTTGGGTAAAGTC  
 306 CTCCTCGACCGCTTGGGGcCCAAATGATCCCCGGCGGAGGTCCCGCAATTGGGTAAAGTC

69/89

FIGURE 6H

139	24	306	CTCTCGGCCCATCTTGGGGCCCAAATGATCCCCGGCGGAGaTCGCGCAATCTGGGTAAGGTC
142	25	306	ATCTCGGCCCATCTTGGGGCCCAAATGATCCCCGGCGTAGGTCCCGCAATCTGGGTAAGGTC
141	21	306	tTCaAGGCCgTCTTGGGGCCcAATGATCCCCGGCGTAGGTCCCGCAATCTGGGTAAGGTC
139-145	consensus		CTCTcGgCCgTCTTGGGGCCcAATGATCCCCGGCGgAGgTCcCGcAAAtTGGGTAAGgTC
SEQ ID NO:	ISOLATE		
145	DK13	367	ATCGATACcCTAACTTGGGGcTTGCGCCGAcCTCATGGGATACATCCCGgTCGTAGGCGCCCC
143	26	367	ATCGATACcCTAACTTGGGGcTTGCGCCGAcCTCATGGGATACATCCCGCTCGTAGGCGCCCC
144	27	367	ATCGATACcCTAACTTGGGGcTTGCGCCGAcCTCATGGGATACATCCCGCTCGTAGGCGCCCC
140	28	367	ATCGATACcCTAACTTGGGGcTTGCGCCGAcCTCATGGGATACATCCCGCTCGTAGGCGCCCC
139	24	367	ATCGATACcCTGACGTGCGGCTTCGCGGACCTCATGGGATACATCCCGaTCGTGGGCGCCCC
142	25	367	ATCGATACcCTGACGTGCGGCTTCGCGGACCTCATGGGATACATTCCGCTCGTcGGGCGCCC
141	21	367	ATCGATACcCTGACGTGCGGCTTCGCGGACCTCATGGGATACATTCCGCTCGTaGGGCGCCC
139-145	consensus		ATCGATACcCT-ACgTGcGGcTTcGCGGAcCTCATGGGATACATcCCGcTCGTaGGGCGCCC
SEQ ID NO:	ISOLATE		
145	DK13	428	CCGTGGGtGGCGTCGCCAGaGCCCTGGCGCATGGcGTcAGGctTcTGGAGGACGGGgTCAA
143	26	428	CCGTGGGcGGCGTCGCCAGGGCCCTGGCaCATGGtGTTAGGGCTgTGGAGGACGGGATCAA
144	27	428	CCGTGGGcGGCGTCGCCAGGGCCCTaGCGCATGGCGTTAGGGCTcTGGAGGACGGGATCAA
140	28	428	CaGTaGgAGGGCGTCGCCAGaGCCCTGGCGCATGGCGTCAGGGCTGTGGAGGACGGGATCAA
139	24	428	CcGTgGGgGGCGTCGCCAGGGCCcTGGCGCATGGCGTCAGGGCTGTGGAGGACGGGATCAA
142	25	428	CaGTaGGTGGCGTCGCCAGGGCCcTGGCGCATGGCGTCAGGGCCcTGGAGGACGGaATCAA
141	21	428	CtGTgGGTGGCGTCGCCAGGGCCcTGGCGCATGGCGTCAGGGCCgTGGAGGACGGaATCAA
139-145	consensus		CcGTgGGtGGCGTCGCCAGgGCCcTgGCGCATGGcGTcAGGgctgTGGAGGACGGgATcAA
SEQ ID NO:	ISOLATE		
145	DK13	489	TTATGCAACAGGGAATCTTCCCGGTGCTCTTCTCTATCTTCTCTTGGCACTgCTcTCG
143	26	489	TTATGCAACAGGGAATCTTCCCGGTGCTCTTCTCTATCTTCTCTTGGCACTTCTTTCG
144	27	489	TTATGCAACAGGGAACCTTCCCGGTGCTCTTCTCTATCTTCTCTTGGCACTTCTTTCG
140	28	489	CTATGCAACAGGGAACCTTCCcGGTGTCTCTTCTCTATCTTCTCTTGGCACTTCTcTCG
139	24	489	CTATGCAACAGGGAATCTTCCcGGTGTCTCTTCTCTATCTTCTCTTGGCACTTCTcTCG
142	25	489	CTATGCAACAGGGAATCTTCTCGGTGCTCCTTCTCTATCTTCTCTTGGCACTTCTcTCG
141	21	489	CTAcGCAACAGGGAACCTTCTCGGTGCTCCTTCTCTATCTTCTCTTGGCACTTCTcTCG
139-145	consensus		cTATGCAACAGGGAATCTTCCcGGTGTCTCCTTCTCTATCTTCTCTcTgGCACTTcTcTCG
SEQ ID NO:	ISOLATE		
145	DK13	550	TGCCTgACTgTTCCTCCgCtTCGGCC
143	26	550	TGCCTaACTgTTCCTCCaCCTCGGCC
144	27	550	TGCCTgACTgTTCCTCCGCTTCGGCC
140	28	550	TGCCTaACcGTcCCAGCGTcGCT
139	24	550	TGCCTcACTGTcCCAGCGTcGCT
142	25	550	TGCcTGACAACACCgGCATCgGCT
141	21	550	TGCcTGACAACACCgGCATCgGcC
139-145	consensus		TGCcTgACTgttCC-gC-TCgGcC

70/89

FIGURE 6I

SEQ ID NO:	ISOLATE	
153	SA11	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
152	SA6	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
146	SA4	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
147	SA5	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
148	SA7	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
149	SA1	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
150	SA3	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
151	SA13	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGCGCCACAGG
146-153	consensus	ATGAGCACGAATCCTAAACCTCAAAGAAAAACCcAAAGAAACACCAACCGcCGCCACAGG
153	SA11	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
152	SA6	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
146	SA4	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
147	SA5	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
148	SA7	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
149	SA1	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
150	SA3	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
151	SA13	62 ACGTCAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTTACTTGTGCGCGCAGGGG
146-153	consensus	ACGTcAAGTTCCCGGGCGGTGGTCAGATCGTTGGTGGAGTTtTACTTGTGCGCGCAGGGG
153	SA11	123 CCCTaGgtTGGGTGTGCGCGGACTCGGAAGACTTCaGAACGGTCGCAACCCCGTGGgCGG
152	SA6	123 CCCTcGtATGGGTGTGCGCGGACTCGGAAGACTTCgGAACGGTCGCAACCCCGTGGaCGG
146	SA4	123 CCCTAGgtTGGGTGTGCGCGGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGGCGG
147	SA5	123 CCCTAGaTTGGGTGTGCGCGGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGGCGG
148	SA7	123 CCCTAGGTtGGGTGTGCGCGGACTCGGAAGACTTCaGAACGGTCGCAACCCCGTGGgCGG
149	SA1	123 CCCCAGGTtGGGTGTGCGCGGACTCGGAAGACTTCgGAACGGTCGCAACCCCGTGGGCGG
150	SA3	123 CCCCAGGTtGGGTGTGCGCGGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGGCGG
151	SA13	123 CCCCAGGTtGGGTGTGCGCGGACTCGGAAGACTTCAGAACGGTCGCAACCCCGTGGGCGG
146-153	consensus	CCCTaGgtTGGGTGTGCGCGGgACTCGGAAGACTTCaGAACGGTCGCAACCCCGTGGgCGG
153	SA11	184 CGTCAGCCTATTCCCAAGGCGCGCCAaCCaCGGGcCGGTCTCTGGGGTCAACCCGGGTACC
152	SA6	184 CGTCAGCCTATTCCCAAGGCGCGCCAAtCCgCGGGtCGGTCTCTGGGGTCAACCCGGGTACC
146	SA4	184 CGCCAGCCTATTCCCAAGGCGCGCCAACCCACGGGCGGTCTCTGGGGTCAACCCGGGTACC
147	SA5	184 CGCCAGCCTATTCCCAAGGCGCGCCAACCCACGGGCGGTCTCTGGGGTCAACCCGGGTACC
148	SA7	184 CGCCAGCCTATTCCCAAGGCGCGCCAACCCACGGGCGGTCTCTGGGGTCAACCCGGGTACC
149	SA1	184 CGCCAGCCTATTCCCAAGGCGCGCCAACCCACGGGCGGTCTCTGGGGTCAACCCGGGTACC
150	SA3	184 CGCCAGCCTATTCCCAAGGcCGCCAGCCcACGGGCGGTCTCTGGGGTCAACCCGGGTACC
151	SA13	184 CGcCAGCCTATeCCCAAGGcGCGCCAGCCcACGGGCGGTCTCTGGGGTCAACCCGGGTACC
146-153	consensus	CGcCAGCCTATtCCCAAGGCGCGCCAaCCaCGGGcCGGTCTCTGGGGTCAACCCGGGTACC
153	SA11	245 CTTGGCCCCtTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGtTGCTCTCCCCcCGAGG
152	SA6	245 CTTGGCCCCtTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTtTGCTCTCCCCcCGAGG
146	SA4	245 CTTGGCCCCtTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTtTGCTCTCCCCcCGAGG
147	SA5	245 CTTGGCCCCtTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTtTGCTCTCCCCcCGAGG
148	SA7	245 CTTGGCCCCtTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTtTGCTCTCCCCcCGAGG
149	SA1	245 CTTGGCCCCtTTTACGCCAATGAGGGCCTCGGGTGGGCAGGGTGGTtTGCTCTCCCCcCGAGG
150	SA3	245 CTTGGCCCCtTTTACGCCAATGAGGGCCTCGaGTGGGCAGGGTGGTtTGCTCTCCCCcCGAGG
151	SA13	245 CTTGGCCCCtTTTaTGCCAATGAGGGCCTCGgGTGGGCAGGGTGGTtTGCTCTCCCCcCGAGG
146-153	consensus	CTTGGCCCCcTTTaTGCCAATGAGGGCCTCGgGTGGGCAGGGTGGtTGCTCTCCCCcCGAGG

71/89

FIGURE 61

SEQ ID NO: ISOLATE

153 SA11  
152 SA6  
146 SA4  
147 SA5  
148 SA7  
149 SA1  
150 SA3  
151 SA13

146-153 consensus

306 CTCTCGGCCTAACTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC  
306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC  
306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC  
306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC  
306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC  
306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC  
306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC  
306 CTCTCGGCCTAATTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC

CTCTCGGCCTAACTGGGGCCCCAATGACCCCGGCGAAATCGCGCAATTGGGGTAAGGTC

SEQ ID NO: ISOLATE

153 SA11  
152 SA6  
146 SA4  
147 SA5  
148 SA7  
149 SA1  
150 SA3  
151 SA13

146-153 consensus

367 ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC  
367 ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC  
367 ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC  
367 ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC  
367 ATCGAACCCTAACATGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC  
367 ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC  
367 ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC  
367 ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC

ATCGATACCCTAACGTGCGGATTGCGCGACCTCATGGGGTACATCCCGCTCGTAGGCGGGC

SEQ ID NO: ISOLATE

153 SA11  
152 SA6  
146 SA4  
147 SA5  
148 SA7  
149 SA1  
150 SA3  
151 SA13

146-153 consensus

428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGAGCTCTTGAGGACGGGGTAAA  
428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTCTTGAGGACGGGGTAAA  
428 CCGTTGGGGGCGTCGCAAGGGCCCTCGCACATGGTGTGAGGGTCTTGAGGACGGGGTAAA  
428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTCTTGAGGACGGGGTAAA  
428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTCTTGAGGACGGGGTAAA  
428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTCTTGAGGACGGGGTAAA  
428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTCTTGAGGACGGGGTAAA  
428 CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGGGTCTTGAGGACGGGGTAAA

CCGTTGGGGGCGTCGCAAGGGCTCTCGCACACGGTGTGAGAGGCTCTTGAGGACGGGGTAAA

SEQ ID NO: ISOLATE

153 SA11  
152 SA6  
146 SA4  
147 SA5  
148 SA7  
149 SA1  
150 SA3  
151 SA13

146-153 consensus

489 CTATGCAACAGGGAATCTCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG  
489 CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG  
489 CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG  
489 CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG  
489 CTACGCAACAGGGAATCTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG  
489 CTACGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG  
489 CTACGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG  
489 CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG

CTATGCAACAGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTTATCCTTGCACTTCTCTCG

SEQ ID NO: ISOLATE

153 SA11  
152 SA6  
146 SA4  
147 SA5  
148 SA7  
149 SA1  
150 SA3  
151 SA13

146-153 consensus

550 TGCTTgACCGTCCCgGCCaCTGCA  
550 TGCCTaACCGTCCCtGCCTCTGCA  
550 TGCTTGACCGTCCCgGCCTCTGCA  
550 TGCTTGACCGTCCCAGCCTCTGCA  
550 TGCCTGACCGTCCCAGCCTCgGCA  
550 TGCTTGATCaTCCCGGCCTCTGCA  
550 TGCCTGACCGTCCCgGCCTCTGCA  
550 TGCCTGACTGTCCCGaCCTCTGCC

TGCTTgACCGTCCCgGCCTCTGCA

72/89

FIGURE 6J

SEQ ID NO:	Genotype	
103-154	cons.	
103-124	1	1 ATGAGCACgaaTCCTAAACCTCAAAGAAaAACCAaAcGtAAcACcAACCGcCGCCcAcagG
125-134	2	1 ATGAGCACAAaTCCtAAACCTCAAAGAAAAACCAaAAGAAaACaAACCGcCGCCcACagG
135-138	3	1 ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAaAAGAAaACCACTTCGTCGCCCACAGG
139-145	4	1 ATGAGCACgAATCCTAAACCTCAAAGAAAAACCAACGTAACACCAACCGcCGCCcATGG
146-153	5	1 ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAaAAGAAaACCAACCGcCGCCcACAGG
154	6	1 ATGAGCACACTTCCAAARCCCAAGAAAAACCAaAAGAAaACCAACCGTCGCCCACCG
103-124	1	62 ACGTcAAGTTCCCGGGcGGtGGtCAGATCGTtGGtGGAGTtTAcTGTtTGCCGCGCAGGGG
125-134	2	62 ACGTcAAGTTCCCGGGcGGcGGcCAGATCGTtGGCGGAGTtTACTTGTtTGCCGCGCAGGGG
135-138	3	62 ACGTcAAGTTCCCGGGTGGCGGACAGATCGTtGGTGGAGTtTACGTTtTGCCGCGCAGGGG
139-145	4	62 AcGTaAAGTTCCCGGGtGGtGGcCAGATCGTtGGCGGAGTtTACTTGTtTGCCGCGCAGGGG
146-153	5	62 ACGTcAAGTTCCCGGGcGGTGGTcAGATCGTtGGTGGAGTtTACTTGTtTGCCGCGCAGGGG
154	6	62 ACGTCAAGTTCCCGGGTGGCGGTcAGATCGTtGGCGGAGTtTACTTGTtTGCCGCGCAGGGG
103-124	1	123 CCCcaGgtTGGGTGTGCGCgCgaCtaGgAAgaCTTCcGAGCGgTCgCAACCTcGtGGaaGg
125-134	2	123 CCCcaGgtTGGGTGTGCGCGCaCtAGGAAGACTTCcGAGCGgTCgCAACCTCGtGGaaGg
135-138	3	123 CCCcAGgTTGGGTGTGCGCGcGACaAGGAagaCTTCcGAGCGaTCCcAGCCgCGTGGgAGg
139-145	4	123 CCCACGATTGGGTGTGCGCGCGACGCGTAAaACTTCTGAACGGTcCaCAGCCTCGCGGACGa
146-153	5	123 CCCcaGgtTGGGTGTGCGCgCgaCTcGgAAGACTTCGAGCGGTcGCAACCTCGtGGcAGg
154	6	123 CCCtaGgtTGGGTGTGCGCGcACTCGGAAGACTTCaGAACGGTTCGCAACCCCGTGGgCGG
123		123 CCCCCGTTGGGTGTGCGCGCAGCAGAAAGACTTCCGAGCGATCCcAGCCcAGAGGCAGG
103-124	1	184 CGaCagCCtATcCCcaAgGctCGcCggccccagGGcaggtcCTGGGctcagCCcGGgtAcc
125-134	2	184 CGaCAaCCTATCCCCAAGGctCGcCggCCCCGAGGgcAGGgCTGGGctCAGCCcGGgtAcc
135-138	3	184 CGCCAGCCCCATCCcAAAGATCGGCGCtCCACTGGCAAGtCCTGGGGAaaCCaGGATATc
139-145	4	184 CGACAGCCTATCCCAAGGCGCGTCGGAGCGAAGGCGGtCCTGGGCTCAGCCcGGGTACC
146-153	5	184 CGtCagCCTATCCCAAGGCaCGtCggtccGAGGgcAGgTCTGGGctCAAGCCcGGGTACC
154	6	184 CGcCAGCCTATtCCCAAGGCGCGCCaaccCaCGGGcCGGTCTCGGGTCAAGCCcGGGTACC
184		184 CGCCAACCTATACCAAGGCGCGCCAGCCCCAGGGCAGGCACTGGGCTCAGCCCGGATACC
103-124	1	245 CtTGGCCcctTcTATGgcaAtGagGGcttcGggTGGGcAGGaTGGcTcTgTCCcCCcCgcGG
125-134	2	245 CtTGGCCCCCTCTATGgCaAtGAGGGcttgGGgTGGGcAGGaTGGCTCCTGTcACCCcgtGG
135-138	3	245 CtTGGCCcCTgTATGGgAAAGAGGGcctCGGctTGGGcAGGtTGGCTCCTGTCCCCCGcGG
139-145	4	245 CTGGCCCCCTCTATGgTAACgAGGGCTGCGGGTGGGcAGGgTGGCTCCTGTCCCCACGCGG
146-153	5	245 CtTGGCCtCtTtAcGGcAAtGagGGctGcGGGTGGGcAGGgTGGCTCtGTCCcCCcCGcGG
154	6	245 CTGGCCcctTTTAcGCCAATGAGGGCCTCGgGTGGGcAGGGTGGtTGCTCTCCCCcGAGG
245		245 CTGGCCCTCTTTATGGAACGAGGGCTGTGGGTGGGcAGGtTGGCTCCTGTCCCCCGCGG
103-124	1	306 cTctcggCctagtTGGGGcCccActGACCCCCGGCgtaggTCgCGcAAAtTGGGtAagGTC
125-134	2	306 cTctCGGCCTAGtTGGGGCCCCcAaGACCCCCGGCGtAGGTcCGGtAAAtTGGGtAAGGTC
135-138	3	306 tTctCgtCCtcttTGGGGCCCCActGACCCCCGGCAtAgATCgCGcAAAtTGGGtAagGTC
139-145	4	306 CTCCCCGTCCATCTTGGGGCCCAaAcGACCCCCGGCGgaGGTCCCGCAATTGGGTAAaGTC
146-153	5	306 cTctcGgCCgTCTTGGGGcCcaAATGATCCCCGGCGgAGgTCcCGcAAAtTGGGtAAGGTC
154	6	306 CTCTCGGCCTAAtTGGGGCCCCAAtGACCCCCGGCGGAaAaTCGCGCAAtTGGGtAAGGTC
306		306 CTCCCCGCCACATTGGGGCCCCAATGACCCCCGGCGTCGATCCCGGAATTGGGTAAAGTC

73/89

FIGURE 6J

SEO ID NO: Genotype  
103-154 cons.

103-124 1  
125-134 2  
135-138 3  
139-145 4  
146-153 5  
154 6

367 ATCGAtACcctACgTGcgGctTcGCCGAcCTCATGGGgTACaTcCCGcTCGTcGGcGccC  
367 ATCGAtACCCCTcACaTGCGGCTTcGCCGACCTCATGGGgTACATcCCGCTCGTCGGcGccc  
367 ATCGAtACCCCTaACgTGcgGcttTTGCCGACCTCATGGGgTACaTcCCcGTCTGtGGCGccc  
367 ATCGAtACCCCTtACGTGCGGaTTCCGCCGACCTCATGGGgTACATCCCGCTCGTCGGCGCTC  
367 ATCGAtACcCTgACgTGcGGcTTcGCCGAcCTCATGGGgTACATcCCGcTCGTaGGCGCCC  
367 ATCGAtACCCCTaACgTGCGGATTCCGCCGAcCTCATGGGgTACATCCCGCTCGTAGGCGGCC  
367 ATCGAtACCCCTAACGTGTGGGTTCGCCGATCTCATGGGgTACATTCCCGCTCGTGGCGCGC

SEO ID NO: Genotype  
103-154 cons.

103-124 1  
125-134 2  
135-138 3  
139-145 4  
146-153 5  
154 6

428 CcgTaGGgGGcGtcGCCaggGCCcTgGCgCatGGcGTcaGgggttcTgGagGACGGggTgAA  
428 CccTaGGgGGcGtcGCCAGgGCCcTGGCgCatGGcGTCCGgGtTcTGGagGACGGCGTGAA  
428 CggTtGGaGGcGTcGCCAGAGCtCTgGCaCatGGtGTgAGgGTcCTGGagGACGGgaTaaa  
428 CcGTAGGaGGCGTCGCAAGAGCCCTCGCGCATGGCGTGAGGGCCCTTGAAGACGGGATAAA  
428 CcGTgGGtGGCGTCGCCAGgGCCcTgGCgCATGGcGTcAGGgctgTGGAGGACGGgaTcaa  
428 CCGTTGGGGCGTCGCAAGGGcTCTcGCACaCGGTGTGAGgGtTCTTGAGGACGGGGTAAA  
428 CTTTGGGCGCGTCGCGGCTCGCCTCGCACATGGCGTGAGGGCAATCGAGGACGGGATCAA

SEO ID NO: Genotype  
103-154 cons.

103-124 1  
125-134 2  
135-138 3  
139-145 4  
146-153 5  
154 6

489 cTatGCAACaGGgAAttTgCCcGGTTGCTcTtTtCTcTATcTTcTcTgGctcTgTgTCC  
489 cTatGCAACAGGGAAAtcTgCCcGGTTGCTcTtTtCTcTATcTTCTcTtTgGcttTGcTgTCC  
489 tTatGCAACaGGgAAttTgCCtGGTTGCTcTtTtCTcTATcTTcTtTgCtTGccTtCTGTCc  
489 TTTcGCAACAGGGAACTTGCCCGTTGCTcTtTtCTcTATcTTCTcTtTgCtCTGTTCTCTc  
489 cTatGCAACAGGGAAAtCTTCCcGGTTGCTcTtTtCTcTATcTTcTtTgGCACTtCTcTCG  
489 cTatGCAACaGGgAAttTgCCCGTTGCTcTtTtCTcTATcTTTtTCCTTGCACTTCTcTCg  
489 TTATGCAACAGGGAAATCTCCCGTTGCTcTtTtCTcTATcTTCTTtTGGCACTACTCTCG

SEO ID NO: Genotype  
103-154 cons.

103-124 1  
125-134 2  
135-138 3  
139-145 4  
146-153 5  
154 6

550 TGcctgaccgtcCCagcttCtgcct  
550 TGcttTgACcatcCCaGctTCCGCT  
550 TGCatCaCagtgCCaGtgTCTGCT  
550 TGCTTAATTcATCCaGCAGCTAGT  
550 TGCcTgACTgttCCagCgTCgGCc  
550 TGccTgAccgTCCCGgCCtCtGCa  
550 TGCCTCACaACGCCAGCTTCGGCT



**FIGURE 6X**

Genotype	10	20	30	40	50	60	70	80	90	100	110	
ATGGCAGggtTCTAACTCCAAAGAGAAACCCGaaACCGTAACACCGCCGCCCGAGCGTCAAGTCCCGCGCGGTGGTCACAGTGGTGGCGAGTTACTTGTGTCCGCG	10	20	30	40	50	60	70	80	90	100	110	
1/1a	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
103-108	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
109-124	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
111/1b	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
111/2a	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
125-128	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
129-133	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
133a	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
133b	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
133-135-136	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
135-139	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
141	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
141b	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
143	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
143c	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
145	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
145d	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
146	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
146d	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
4f	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
5a	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
146-153	...	GAA	T	...	A	AA	C	T	C	CA	GT	...
154	...	GAA	T	...	A	AA	C	T	C	CA	GT	...

SEC ID NO:	Genotype
103-108	TA-AT
109-124	GA-CA-G-GA
125-128	GA-TA-G-GA
129-133	GA-CA-G-GA
136	GA-CA-G-GA
137	GA-AT
138	GA-AT
139	GA-AT
140	GA-AT
141	GA-AT
142	GA-AT
143	GA-AT
144	GA-AT
145	GA-AT
146	GA-AT
147	GA-AT
148	GA-AT
149	GA-AT
150	GA-AT
151	GA-AT
152	GA-AT
153	GA-AT
154	GA-AT
155	GA-AT
156	GA-AT
157	GA-AT
158	GA-AT
159	GA-AT
160	GA-AT
161	GA-AT
162	GA-AT
163	GA-AT
164	GA-AT
165	GA-AT
166	GA-AT
167	GA-AT
168	GA-AT
169	GA-AT
170	GA-AT
171	GA-AT
172	GA-AT
173	GA-AT
174	GA-AT
175	GA-AT
176	GA-AT
177	GA-AT
178	GA-AT
179	GA-AT
180	GA-AT
181	GA-AT
182	GA-AT
183	GA-AT
184	GA-AT
185	GA-AT
186	GA-AT
187	GA-AT
188	GA-AT
189	GA-AT
190	GA-AT
191	GA-AT
192	GA-AT
193	GA-AT
194	GA-AT
195	GA-AT
196	GA-AT
197	GA-AT
198	GA-AT
199	GA-AT
200	GA-AT
201	GA-AT
202	GA-AT
203	GA-AT
204	GA-AT
205	GA-AT
206	GA-AT
207	GA-AT
208	GA-AT
209	GA-AT
210	GA-AT
211	GA-AT
212	GA-AT
213	GA-AT
214	GA-AT
215	GA-AT
216	GA-AT
217	GA-AT
218	GA-AT
219	GA-AT
220	GA-AT
221	GA-AT
222	GA-AT
223	GA-AT
224	GA-AT
225	GA-AT
226	GA-AT
227	GA-AT
228	GA-AT
229	GA-AT
230	GA-AT
231	GA-AT
232	GA-AT
233	GA-AT
234	GA-AT
235	GA-AT
236	GA-AT
237	GA-AT
238	GA-AT
239	GA-AT
240	GA-AT
241	GA-AT
242	GA-AT
243	GA-AT
244	GA-AT
245	GA-AT
246	GA-AT
247	GA-AT
248	GA-AT
249	GA-AT
250	GA-AT
251	GA-AT
252	GA-AT
253	GA-AT
254	GA-AT
255	GA-AT
256	GA-AT
257	GA-AT
258	GA-AT
259	GA-AT
260	GA-AT
261	GA-AT
262	GA-AT
263	GA-AT
264	GA-AT
265	GA-AT
266	GA-AT
267	GA-AT
268	GA-AT
269	GA-AT
270	GA-AT
271	GA-AT
272	GA-AT
273	GA-AT
274	GA-AT
275	GA-AT
276	GA-AT
277	GA-AT
278	GA-AT
279	GA-AT
280	GA-AT
281	GA-AT
282	GA-AT
283	GA-AT
284	GA-AT
285	GA-AT
286	GA-AT
287	GA-AT
288	GA-AT
289	GA-AT
290	GA-AT
291	GA-AT
292	GA-AT
293	GA-AT
294	GA-AT
295	GA-AT
296	GA-AT
297	GA-AT
298	GA-AT
299	GA-AT
300	GA-AT

SEQ ID NO:	Genotype	240	250	260	270	280	290	300	310	320	330	340
103-108	1/1a	TCAG-C-GT-C-T	CC-C-T-GCA-T-T	G-C-TGCG-GG	G-A	C-CC-G-T	C-GT-C	TCGG-TAGC	C-CG-CB	C	GTAGG	
109-124	1/1b	TCAG-C-GT-C-T	CC-C-T-GCA-T	G-CATCG-GG	B-A	C-CC-G-B	C-GC-C	TCGG-TAGT	C-CG-CG	C	GTAGG	
125-128	1/1/2a	AAAG-A-AT-C-T	CC-C-T-GGA-T	G-ATCG-GG	B-A	C-CC-G-C	C-GA-T	CCGTCTCT	C-CC-AT	C	ATAGG	
129-133	1/2b	AAAG-A-AT-T	CC-B-T-GGA-T	G-CATCG-GC	A-T	C-CC-G-C	C-GC-T	TCGT-TACT	C-CC-CT	C	ATAGA	
134	2c	ACGG-A-AT-T	CC-B-T-GGA-T	G-CATCG-GC	A-G	C-CC-G-C	C-GC-T	TCGT-TTCA	C-CC-CT	C	ATAAA	
135-138	3a	TCAG-C-GT-C-T	CC-C-T-GTA-S	G-CATCG-GG	A-B	C-CC-G-T	A-GC-C	TCGG-TACT	C-CA-AC	C	GGAGG	
139	4a	TCAG-C-GT-C-T	CC-C-T-GCA-T	G-CATCG-GG	A-G	C-CC-G-T	A-GC-C	TCGG-TACT	C-CA-AT	C	GGAGA	
140	4b	TCAG-C-GT-C-T	CC-T-CGCA-T	G-CATCG-GG	A-G	C-CC-G-T	A-GC-G	TCGG-TACT	C-CC-AT	C	GTAGG	
141	4c	TCAG-C-GT-C-T	CC-T-CGTA-T	G-CATCG-GG	B-A	C-CC-G-A	C-GT-C	TCGA-GTCT	C-CC-AT	C	GTAGG	
142	4d	TCAG-T-GT-T	CC-T-CGCA-T	G-CATCG-GG	A-G	C-CC-G-A	C-GT-C	TCGG-TACT	C-CG-AT	C	GTAGG	
143	4e	TCAG-C-GT-C-T	CC-T-CGCA-T	G-CATCG-GG	A-G	C-CC-G-C	C-GC-A	TCGG-TACT	C-AA-AT	C	GTAGG	
144	4f	TCAG-C-GT-C-T	CC-T-CGTA-T	G-CATCG-GG	A-T	C-CC-G-C	C-GC-C	TCGG-TACT	C-CA-AT	C	GTAGG	
145	5a	TCAG-C-GT-C-T	CC-C-T-GCA-T	G-CATCG-GG	A-G	C-CC-G-C	C-GA-C	TCGG-TACT	C-CC-AT	C	GGAAB	
146-153	5b	TCAG-C-GT-C-T	CC-T-T-GAA-C	G-CATCG-GG	A-T	C-CC-G-C	C-GC-C	TCGG-TACT	C-CC-AT	C	GTAGA	
154	6a	TCAG-C-GT-C-T	CC-T-T-GAA-C	G-CATCG-GG	A-T	C-CC-G-C	C-GC-C	TCGG-TACT	C-CC-AT	C	GTAGA	



76/89

FIGURE 7A

SEQ ID NO: ISOLATE

156 US11  
157 S14  
158 SW1  
159 S18  
160 DR4  
155 DK7

155-160 consensus

1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR

SEQ ID NO: ISOLATE

156 US11  
157 S14  
158 SW1  
159 S18  
160 DR4  
155 DK7

155-160 consensus

62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV  
62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV  
62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV  
62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV  
62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV  
62 RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV

RQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPTDPRRRSRNLGKV

SEQ ID NO: ISOLATE

156 US11  
157 S14  
158 SW1  
159 S18  
160 DR4  
155 DK7

155-160 consensus

123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS

IDTLTCGFADLMGYIPLVGAPLGGAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS

SEQ ID NO: ISOLATE

156 US11  
157 S14  
158 SW1  
159 S18  
160 DR4  
155 DK7

155-160 consensus

184 CLTVPASA  
184 CLTVPASA  
184 CLTVPASA  
184 CLTVPASA  
184 CLTVPASA  
184 CLTVPASA

CLTVPASA

77/89

**FIGURE 7B**

[illegible]

**MST<sub>n</sub>PKPORKTKRNT<sub>n</sub>RRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSESRQPRGR**

[illegible]

ROIPKARrPEGRaWAOPGyPWPLYgnEG-GWAGWLLSPtGSRPbWGPTcDPRRRSRNLGKV

[illegible]

**IDTLTCGFADLMGYIPLVGaPLGGaARALAHGVRVIEDGVNYATGNIPGCsFSIFLLALLS**

184 CLT1PASA  
184 CLTvPASA  
184 CLTIPASA

78/89

## FIGURE 7B

171	S9	184	CLTIPASA
163	D1	184	CLTIPASA
165	P10	184	CLTIPASA
169	IND3	184	CLTIPASA
164	US6	184	CLTIPASA
166	DK1	184	CLTIPASA
167	T10	184	CLTIPASA
168	SW2	184	CLTIPASA
161	SA10	184	CLTIPASA
174	HK4	184	CLTIPASA
172	HK3	184	CLTIPASA
176	T3	184	CLTIPASA
173	HK5	184	CLTIPASA
161-176	consensus		CLTIPASA



80/89

FIGURE 7C

159	S18	123	IDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
160	DR4	123	IDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
155	DK7	123	IDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
170	IND8	123	IDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
162	S45	123	IDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
171	S9	123	IDTLTCGFADLMGYIPLVGAPLGGAAARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS
175	P8	123	IDTLTCGFADLMGYIPLVGgPLGGvARALAHGVRVvEDGVNYATGNLPGCSFSIFLLALLS
155-176	consensus		IDTLTCGFADLMGYIPLVGaPLGGaARALAHGVRVIEDGVNYATGNIPGCsFSIFLLALLS
SEQ ID NO:	ISOLATE		
173	HK5	184	CLTtPvSA
176	T3	184	CLTtPASA
172	HK3	184	CLTtPASA
174	HK4	184	CLTtPASA
161	SA10	184	CLTtPASA
168	SW2	184	CLTtPASA
167	T10	184	CLTtPASA
166	DK1	184	CLTtPASA
164	US6	184	CLTtPASA
169	IND3	184	CLTtPASA
165	P10	184	CLTtPASA
163	D1	184	CLTtPASA
156	US11	184	CLTVPASA
157	S14	184	CLTVPASA
158	SW1	184	CLTVPASA
159	S18	184	CLTVPASA
160	DR4	184	CLTVPASA
155	DK7	184	CLTVPASA
170	IND8	184	CLTVPASA
162	S45	184	CLTtPASA
171	S9	184	CLTtPASA
175	P8	184	CLTtPASA
155-176	consensus		CLTtPASA

81/89

FIGURE 7D

SEQ ID NO: ISOLATE

179 T9  
178 US10  
180 T2  
177 T4

177-180 consensus

1 MSTNPKPQKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
1 MSTNPKPQKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR

SEQ ID NO: ISOLATE

179 T9  
178 US10  
180 T2  
177 T4

177-180 consensus

62 RQIPKORRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPdDPRHRSRNVGKV  
62 RQIPKORRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPdDPRHRSRNVGKV  
62 RQIPKORRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPdDPRHRSRNVGKV  
62 RQIPKORRSTGKSWGKPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPdDPRHRSRNVGKV

SEQ ID NO: ISOLATE

179 T9  
178 US10  
180 T2  
177 T4

177-180 consensus

123 IDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCGFADLMGYIPVVGAPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS  
123 IDTLTCsLADLMGYIPVVGgPLGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFLLALLS

SEQ ID NO: ISOLATE

179 T9  
178 US10  
180 T2  
177 T4

177-180 consensus

184 CITtPaSA  
184 CITiPVSA  
184 CITiPVSA  
184 CITiPVSA

CITiPVSA



82/89

FIGURE 7E

SEQ ID NO: ISOLATE  
 183 DK11  
 184 SW3  
 181 T8  
 182 US1  
 185 DK8

181-185 consensus

1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRtTRKTSERSQPRGR  
 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR  
 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKtSERSQPRGR  
 1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKsSERSQPRGR  
 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKtSERSQPRGR

SEQ ID NO: ISOLATE  
 183 DK11  
 184 SW3  
 181 T8  
 182 US1  
 185 DK8

181-185 consensus

62 RQIPKDRRSTGKsWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHkSRNLGKV  
 62 RQIPKDRRSTGKsWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV  
 62 RQIPKDRRSTGKsWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLGxV  
 62 RQIPKDRRSTGKsWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLGKV  
 62 RQIPKDRRSTGKsWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPTWGPTDPRHRSRNLGKV  
 RQIPKDRRSTGKsWGKPGYPWPLYGNEGCGWAGWLLSPRGSRPtWGPTDPRHrSRNLGkV

SEQ ID NO: ISOLATE  
 183 DK11  
 184 SW3  
 181 T8  
 182 US1  
 185 DK8

181-185 consensus

123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS  
 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS  
 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS  
 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS  
 123 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS  
 IDTITCGFADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS

SEQ ID NO: ISOLATE  
 183 DK11  
 184 SW3  
 181 T8  
 182 US1  
 185 DK8

181-185 consensus

184 CcTVPVSA  
 184 CFTVPVSA  
 184 CFTVPVSA  
 184 CaTVPVSA  
 184 CcTVPVSA

C-TV PVSA

83/89

FIGURE 7F

SEQ ID NO:	ISOLATE	
183	DK11	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
184	SW3	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
181	T8	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
182	US1	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
185	DK8	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
186	S83	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
178	US10	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
180	T2	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
179	T9	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
177	T4	1 MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
177-186	consensus	MSTNPKPQRTKRNTRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
183	DK11	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
184	SW3	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
181	T8	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
182	US1	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
185	DK8	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
186	S83	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
178	US10	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
180	T2	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
179	T9	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
177	T4	62 RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
177-186	consensus	RQPIPKDRRSTGKSWGKPGYPWPLYGNEGCGWAGWLLSPRGSHPNWGPTDPRHRSRNLGKV
183	DK11	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
184	SW3	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
181	T8	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
182	US1	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
185	DK8	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
186	S83	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
178	US10	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
180	T2	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
179	T9	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
177	T4	123 IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
177-186	consensus	IDTTTCGFAADLMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLS
183	DK11	184 CcTVPVSA
184	SW3	184 CcTVPVSA
181	T8	184 CcTVPVSA
182	US1	184 CcTVPVSA
185	DK8	184 CcTVPVSA
186	S83	184 CcTVPVSA
178	US10	184 CcTVPVSA
180	T2	184 CcTVPVSA
179	T9	184 CcTVPVSA
177	T4	184 CcTVPVSA
177-186	consensus	CcTVPVSA

84/89

## FIGURE 7G

<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
189	S2	1 MSTLPKPQRTKRTTIRRPQDVKFPGGGQIVGGVYVLPRRGPRLGVRATRTKTSERSQPRGR
187	HK10	1 MSTLPKPQRTKRTTIRRPQDVKFPGGGQIVGGVYVLPRRGPRLGVRATRTKTSERSQPRGR
190	DK12	1 MSTLPKPQRTKRTTIRRPQDVKFPGGGQIVGGVYVLPRRGPRLGVRATRTKTSERSQPRGR
188	S52	1 MSTLPKPQRTKRTTIRRPQDVKFPGGGQIVGGVYVLPRRGPRLGVRATRTKTSERSQPRGR
187-190	consensus	MSTLPKPQRTKRTTIRRPQDVKFPGGGQIVGGVYVLPRRGPRLGVRATRTKTSERSQPRGR
<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
189	S2	62 RQIPKARRSEGRSWAQP GYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
187	HK10	62 RQIPKARRSEGRSWAQP GYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
190	DK12	62 RQIPKARRSEGRSWAQP GYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
188	S52	62 RQIPKARRSEGRSWAQP GYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
187-190	consensus	RQIPKARRSEGRSWAQP GYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
189	S2	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
187	HK10	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
190	DK12	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
188	S52	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
187-190	consensus	IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINFATGNLPGCSFSIFLLALFS
<u>SEQ ID NO:</u>	<u>ISOLATE</u>	
189	S2	184 CLINHPAAS
187	HK10	184 CLINHPAAS
190	DK12	184 CLINHPAAS
188	S52	184 CLVHPAAS
187-190	consensus	CLINHPAAS

85/89

FIGURE 7H

SEQ ID NO:	ISOLATE	
194	25	1 MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
193	21	1 MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
192	28	1 MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
195	26	1 MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
196	27	1 MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
191	24	1 MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
197	DK13	1 MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
191-197	consensus	MSTNPKPQRTKRTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
194	25	62 RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
193	21	62 RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
192	28	62 RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
195	26	62 RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
196	27	62 RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
191	24	62 RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
197	DK13	62 RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
191-197	consensus	RQPIPKARRSEGRSWAOPGYWPPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKV
194	25	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
193	21	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
192	28	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
195	26	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
196	27	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
191	24	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
197	DK13	123 IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
191-197	consensus	IDTLTCGFADLMGYIPLVGAPVGGVARALAHGVRALEDGINYATGNLPGCSFSIFLLALLS
194	25	184 CLTTPASA
193	21	184 CLTTPASA
192	28	184 CLTVPASA
195	26	184 CLTVPASA
196	27	184 CLTVPASA
191	24	184 CLTVPASA
197	DK13	184 CLTVPASA
191-197	consensus	CLTVPASA

86/89

FIGURE 71

SEQ ID NO:	ISOLATE	
205	SA11	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
202	SA3	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
198	SA4	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
199	SA5	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
200	SA7	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
203	SA13	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
201	SA1	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
204	SA6	1 MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
198-205	consensus	MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGR
SEQ ID NO:	ISOLATE	
205	SA11	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
202	SA3	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
198	SA4	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
199	SA5	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
200	SA7	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
203	SA13	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
201	SA1	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
204	SA6	62 RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
198-205	consensus	RQPIPKAROPTGRSWGQPGYPWPPLYANEGLGWAGWLLSPRGSRPNWGPNDPRRKSRNLGKV
SEQ ID NO:	ISOLATE	
205	SA11	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
202	SA3	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
198	SA4	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
199	SA5	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
200	SA7	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
203	SA13	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
201	SA1	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
204	SA6	123 IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
198-205	consensus	IDTLTCGFADLMGYIPLVGGPVGGVARALAHGVRVLEDGVNYATGNLPGCSFSIFILALLS
SEQ ID NO:	ISOLATE	
205	SA11	184 CLTVPAtA
202	SA3	184 CLTVPASA
198	SA4	184 CLTVPASA
199	SA5	184 CLTVPASA
200	SA7	184 CLTVPASA
203	SA13	184 CLTVPASA
201	SA1	184 CLTVPASA
204	SA6	184 CLTVPASA
198-205	consensus	CLTVPASA

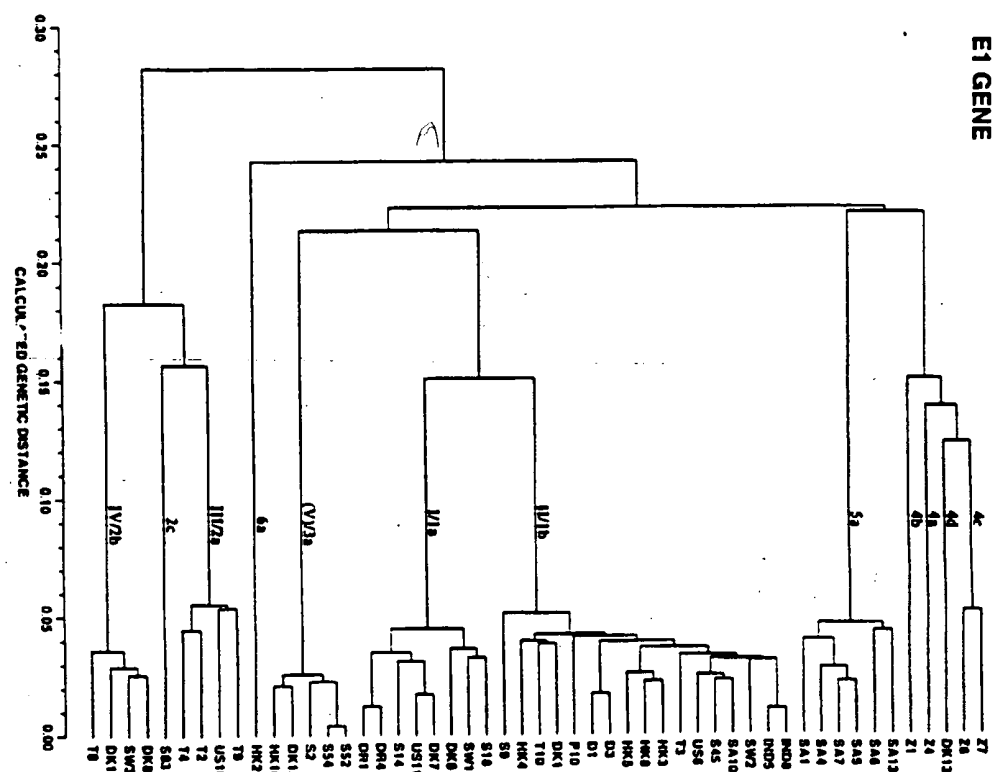
87/89

FIGURE 7J

SEQ. ID. NO.	Genotype		
155-206	cons.	1	MSTnPKPQRKTKRNTnRPQdVKFPGGQIVGGVYLLPRRGPRIGVRatRktSERSQPRGRRQPIPKarPteGzawaqPQyPWPlygnEGCGWAGW
155-176	type 1		MSTnPKPQRKTKRNTnRRPQdVKFPGGQIVGGVYLLPRGPRLGVRatRktSERSQPRGRRQPIPKarPteGzawaqPQyPWPlygnEGCGWAGW
177-186	type 2		MSTnPKPQRKTKRNTnRRPQdVKFPGGQIVGGVYLLPRGPRLGVRatRktSERSQPRGRRQPIPKarPteGzawaqPQyPWPlygnEGCGWAGW
187-190	type 3		MSTnPKPQRKTKRNTnRRPQdVKFPGGQIVGGVYLLPRGPRLGVRatRktSERSQPRGRRQPIPKarPteGzawaqPQyPWPlygnEGCGWAGW
191-197	type 4		MSTnPKPQRKTKRNTnRRPQdVKFPGGQIVGGVYLLPRGPRLGVRatRktSERSQPRGRRQPIPKarPteGzawaqPQyPWPlygnEGCGWAGW
198-205	type 5		MSTnPKPQRKTKRNTnRRPQdVKFPGGQIVGGVYLLPRGPRLGVRatRktSERSQPRGRRQPIPKarPteGzawaqPQyPWPlygnEGCGWAGW
198-206	type 6		MSTnPKPQRKTKRNTnRRPQdVKFPGGQIVGGVYLLPRGPRLGVRatRktSERSQPRGRRQPIPKarPteGzawaqPQyPWPlygnEGCGWAGW
155-206	cons.	97	LLSPtGSRPawGptDPRrrSRNlGkVIDtLTcfADLMGYiPlVGaPlGGVaRALAHGVRvIEDGvNyATGNlPGCsfSIFlLlLlSCLtVPaBa
155-176	type 1		LLSPtGSRPawGptDPRrrSRNlGkVIDtLTcfADLMGYiPlVGaPlGGVaRALAHGVRvIEDGvNyATGNlPGCsfSIFlLlLlSCLtVPaSa
177-186	type 2		LLSPtGSRPawGptDPRrrSRNlGkVIDtLTcfADLMGYiPlVGaPlGGVaRALAHGVRvIEDGvNyATGNlPGCsfSIFlLlLlSCLtVPaSa
187-190	type 3		LLSPtGSRPawGptDPRrrSRNlGkVIDtLTcfADLMGYiPlVGaPlGGVaRALAHGVRvIEDGvNyATGNlPGCsfSIFlLlLlSCLtVPaSa
191-197	type 4		LLSPtGSRPawGptDPRrrSRNlGkVIDtLTcfADLMGYiPlVGaPlGGVaRALAHGVRvIEDGvNyATGNlPGCsfSIFlLlLlSCLtVPaSa
198-205	type 5		LLSPtGSRPawGptDPRrrSRNlGkVIDtLTcfADLMGYiPlVGaPlGGVaRALAHGVRvIEDGvNyATGNlPGCsfSIFlLlLlSCLtVPaSa
198-206	type 6		LLSPtGSRPawGptDPRrrSRNlGkVIDtLTcfADLMGYiPlVGaPlGGVaRALAHGVRvIEDGvNyATGNlPGCsfSIFlLlLlSCLtVPaSa



## FIGURE 8





**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 6 :</b> <b>C12N 15/51, C07K 14/18, G01N 33/53, A61K 39/29, C12Q 1/68, 1/70, C07K 16/10</b>	<b>A3</b>	<b>(11) International Publication Number: WO 96/05315</b> <b>(43) International Publication Date: 22 February 1996 (22.02.96)</b>
<b>(21) International Application Number: PCT/US95/10398</b> <b>(22) International Filing Date: 15 August 1995 (15.08.95)</b>  <b>(30) Priority Data:</b> 08/290,665 15 August 1994 (15.08.94) US  <b>(71) Applicant: THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES, Office of Technology Transfer National Institutes of Health [US/US]; Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).</b>  <b>(72) Inventors: BUKH, Jens; 5805 Sonoma Road, Bethesda, MD 20817 (US). MILLER, Roger, H.; 15504 White Willow Lane, Rockville, MD 20853 (US). PURCELL, Robert, H.; 17517 White Grounds Road, Boyds, MD 20841 (US).</b>  <b>(74) Agent: FEILER, William, S.; Morgan &amp; Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).</b>	<b>(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).</b>  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>  <b>(88) Date of publication of the international search report: 4 April 1996 (04.04.1996)</b>	
<b>(54) Title: NUCLEOTIDE AND AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND CORE GENES OF HEPATITIS C VIRUS</b>  <b>(57) Abstract</b>  The nucleotide and deduced amino acid sequences of cDNAs encoding the envelope (1) genes and core genes of isolates of hepatitis C virus (HCV) are disclosed. The invention relates to the oligonucleotides, peptides and recombinant envelope (1) and core proteins derived from these sequences and their use in diagnostic methods and vaccines.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## INTERNATIONAL SEARCH REPORT

International Application No

PC1/US 95/10398

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/51 C07K14/18  
C12Q1/70 C07K16/10

G01N33/53

A61K39/29

C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K G01N A61K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROC. NATL. ACAD. SCI. USA., vol. 89, no. 11, June 1992 pages 4942-4946, JENS BUKH ET AL. 'Sequence analysis of the 5' noncoding region of hepatitis C virus.'	1,2,21, 31,32,38
Y	see the whole document	5-20, 22-30, 33-37, 39-51, 53,54, 56-59
	---	
	-/--	

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

20 February 1996

Date of mailing of the international search report

01.03.96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

Hix, R

## INTERNATIONAL SEARCH REPORT

Inter national Application No

PCT/US 95/10398

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROC NATL ACAD SCI U S A 90 (17). 1993. 8234-8238. CODEN: PNASA6 ISSN: 0027-8424, September 1993 BUKH J ET AL 'AT LEAST 12 GENOTYPES OF HEPATITIS C VIRUS PREDICTED BY SEQUENCE ANALYSIS OF THE PUTATIVE E1 GENE OF ISOLATES COLLECTED WORLDWIDE.'	1,2,5, 21-23, 26-32,38
Y	see the whole document	5-20,24, 25, 33-37, 39-51, 53,54, 56-59
X	--- JOURNAL OF GENERAL VIROLOGY 75 (5). 1994. 1053-1061. ISSN: 0022-1317, May 1994 SIMMONDS P ET AL 'Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions.'	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
X	--- JOURNAL OF BIOMEDICAL SCIENCE 1 (3). 1994. 158-162. ISSN: 1021-7770, June 1994 KAO J-H ET AL 'Detection of divergent hepatitis C virus envelope sequences.'	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
X	--- BIOCHEM BIOPHYS RES COMMUN 192 (2). 1993. 635-641. CODEN: BBRCA9 ISSN: 0006-291X, 30 April 1993 STUYVER L ET AL 'ANALYSIS OF THE PUTATIVE E1 ENVELOPE AND NS4A EPITOPE REGIONS OF HCV TYPE 3.'	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
	--- -/--	

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 95/10398

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ARCHIVES OF VIROLOGY SUPPLEMENTUM 0 (7). 1993. 27-39. ISSN: 0939-1983, ROGGENDORF M ET AL 'Variability of the envelope regions of HCV in European isolates and its significance for diagnostic tools.'	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
X	--- PROC. NATL. ACAD. SCI. U. S. A. (1992), 89(15), 7144-8 CODEN: PNASA6;ISSN: 0027-8424, 1 August 1992 CHA, T. A. ET AL 'At least five related, but distinct, hepatitis C viral genotypes exist'	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
X	--- BIOCHEM. BIOPHYS. RES. COMMUN. (1994), 199(3), 1474-81 CODEN: BBRC A9;ISSN: 0006-291X, 30 March 1994 LI, JI-SU ET AL 'Identification of the third major genotype of hepatitis C virus in France'	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
X	--- WO,A,94 01778 (CHIRON CORP) 20 January 1994	53,54, 56-58
Y	see the whole document	1,2, 5-51,59
X	--- WO,A,92 19743 (CHIRON CORP) 12 November 1992	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
	--- -/--	

Internal Application No  
PC1/US 95/10398

PC7/US 95/10398

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO,A,92 21759 (PASTEUR INSTITUT) 10 December 1992  see the whole document ---	1,2, 5-51,53, 54,56-59
X	EP,A,0 586 065 (TONEN CORP) 9 March 1994  see the whole document ---	53,54, 56-58
X	PROC. NATL. ACAD. SCI. USA, vol. 89, January 1992 pages 187-191, J. BUKH ET AL 'Importance of primer selection for the detection of hepatitis C virus RNA with the polymerase chain reaction assay' see the whole document ---	21,28
P,X	WO,A,95 01442 (US HEALTH) 12 January 1995  see the whole document ---	1,2, 5-51,53, 54,56-59
P,X	WO,A,94 25601 (INNOGENETICS NV ;MAERTENS GEERT (BE); STUYVER LIEVEN (BE)) 10 November 1994	21,28
Y	see the whole document	1,2, 5-20, 22-27, 29-51, 53,54, 56-59
P,X	WO,A,94 27153 (CHIRON CORP) 24 November 1994	21,28, 53-58
Y	see the whole document	1,2, 5-20, 22-27, 29-51,59
P,X	PROC. NATL. ACAD. SCI. U. S. A. (1994), 91(21), 10134-8 CODEN: PNASA6;ISSN: 0027-8424, 11 October 1994 STUYVER, LIEVEN ET AL 'Classification of hepatitis C viruses based on phylogenetic analysis of the envelope 1 and nonstructural 5B regions and identification of five additional subtypes' see the whole document ---	1,2, 5-51,53, 54,56-59

-/--

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 95/10398

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	JOURNAL OF CLINICAL MICROBIOLOGY 32 (9). 1994. 2280-2284. ISSN: 0095-1137, September 1994 RAVAGGI A ET AL 'Distribution of viral genotypes in Italy determined by hepatitis C virus typing by DNA immunoassay.' see the whole document ---	1,2, 5-51,53, 54,56-59
P,X	SEMINARS IN LIVER DISEASE, vol. 15, no. 1, February 1995 pages 41-63, J. BUKH ET AL. 'Genetic Heterogeneity of hepatitis C virus: Quasispecies and genotypes' see the whole document ---	1,2, 5-51,53, 54,56-59
P,X	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 202, no. 3, 15 August 1994 pages 1308-1314, L. STUYVER 'Cloning and phylogenetic analysis of the core, E2 and NS3/NS4 regions of the hepatitis C virus type 5a+' see the whole document -----	1,2, 5-51,53, 54,56-59

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US95/10398

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 18,45,49  
because they relate to subject matter not required to be searched by this Authority, namely:  
Remark: Although these claims are directed to a method of treatment of (diagnostic method practised on) the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

- 26 subjects

See continuation-sheets PCT/ISA/210

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1,2,5-51,53,54,56-59 (partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.



FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

claims:

1. 1,2,5-51,53,54,56 to 59 (partially):

Genotypes specific peptides from E1 Seq. ID 1-8 and 52-59 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype I/1a.

2. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 103-108 and 155-160 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype I/1a.

3. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 9-25 and 60-76 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype II/1b.

4. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 109-124 and 161-176 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype II/1b.

5. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 26-29 and 77-80 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype III/2a.

6. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 125-128 and 177-180 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype III/2a.

7. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 30-33 and 81-84 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype IV/2b.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

## 8. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 129-133 and 181-185 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype IV/2b.

## 9. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 34 and 85 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype IV/2c.

## 10. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 134 and 186 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype IV/2c.

## 11. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 35-39 and 86-90 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype V/3a.

## 12. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 135-138 and 187-190 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype V/3a.

## 13. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 40 and 91 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4a.

## 14. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 139 and 191 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4a.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

## 15. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 41 and 92 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4b.

## 16. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 141 and 193 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4b.

## 17. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 42-43 and 93-94 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4c.

## 18. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 143-144 and 195-196 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4c.

## 19. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 44 and 95 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4d.

## 20. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 145 and 197 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4d.

## 21. 3-52,55 and 59 (partially):

Genotype specific peptides Core Seq. ID 142 and 194 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4e.

FURTHER INFORMATION CONTINUED FROM PCT/SAJ/210

## 22. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 140 and 192 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 4f.

## 23. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 45-50 and 96-101 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 5a.

## 24. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. 146-153 and 198-205 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 5a.

## 25. 1,2,5-51,53,54,56 to 59 (partially):

Genotype specific peptides from E1 Seq. ID 51 and 102 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 6a.

## 26. 3-52,55 and 59 (partially):

Genotype specific peptides from Core Seq. ID 154 and 206 used in the recombinant protein expression, detection of antibodies against HCV, vaccines and methods of detection using PCR primers and Identification of Genotype 6a.

---

## INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No.

PCT/US 95/10398

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9401778	20-01-94	AU-B- 4662993	31-01-94
		EP-A- 0649537	26-04-95
		FI-A- 950002	27-02-95
		HU-A- 70473	30-10-95
		JP-T- 7509060	05-10-95
		NO-A- 950006	24-02-95
		PL-A- 307178	02-05-95
		SK-A- 495	11-07-95
WO-A-9219743	12-11-92	AU-B- 2155892	21-12-92
		BG-A- 98200	31-01-95
		CZ-A- 9302377	13-04-94
		EP-A- 0585398	09-03-94
		HU-A- 69609	28-09-95
		JP-T- 6508026	14-09-94
		NO-A- 934019	05-11-93
		SK-A- 123293	08-06-94
WO-A-9221759	10-12-92	FR-A- 2677372	11-12-92
		CA-A- 2088666	07-12-92
		EP-A- 0542970	26-05-93
		JP-T- 6500698	27-01-94
EP-A-0586065	09-03-94	CN-A- 1087915	15-06-94
		JP-A- 7133292	23-05-95
WO-A-9501442	12-01-95	AU-B- 7319194	24-01-95
WO-A-9425601	10-11-94	AU-B- 6722294	21-11-94
		CA-A- 2139100	10-11-94
		CN-A- 1108030	06-09-95
		EP-A- 0651807	10-05-95
		FI-A- 946066	23-12-94
		JP-T- 7508423	21-09-95
		NO-A- 944967	21-12-94
WO-A-9427153	24-11-94	NONE	